31 No. 2 (New Series No. 139) December 197

ANALYSIS

006 -- H02965 - 1 - P000438

incomplete
missing issue

Edited by
PETER WINCH

CONTENTS

Wittgenstein, philosophy and logic — ILHAM DILMAN

A model for the religious philosophy of D. Z. Phillips — ALAN BRUNTON

Feature universals and sortal universals — MICHAEL DURRANT

'Exists' as a predicate—a reconsideration — JAMES CHILD *and* FRED I. GOLDBERG

The limits of historical knowledge — BURLEIGH T. WILKINS

Time and Time' again — EDDY M. ZEMACH

ASIL BLACKWELL · BROAD STREET · OXFORD

Five shillings net
(Twenty five pence)

ANALYSIS

Edited by Peter Winch

CALCUTTA UNIVERSITY
CENTRAL LIBRARY

Volume 31

1970–1971

1

BASIL BLACKWELL · BROAD STREET · OXFORD

INDEX

		PAGE
ABBOTT, W. R.	What knowledge is not	143
BELL, J. M.	Opacity and identity	19
BLACK, CAROLYN	Knowledge without belief	152
BRUNTON, ALAN	A model for the religious philosophy of D. Z. Phillips	43
CANDLISH, STEWART	Physiological discoveries: criteria or symptoms?	162
CARGILE, JAMES	On near knowledge	145
CHILD, JAMES *and* GOLDBERG, FRED I.	'Exists' as a predicate—a reconsideration	53
DILMAN, ILHAM	Wittgenstein, philosophy and logic	33
DINELLO, DANIEL	On killing and letting die	83
DURRANT, MICHAEL	Feature universals and sortal universals	49
GARNER, RICHARD T.	Some doubts about illocutionary negation	106
GENOVA, A. C.	"Jonesese" and substitutivity	96
GEORGE, F. H.	Belief statements and their logic	104
GOLDBERG, FRED I.	(*See* Child, James)	53
HELM, PAUL	Pretending and intending	127
HERZBERGER, HANS G.	The expressive capacity of non-translational languages	186
JONES, O. R.	On truth—a reply to C. J. F. Williams	24
LYCAN, WILLIAM G.	Williams and Stroud on Shoemaker's logic	159
MACKAY, ALFRED F.	The principle of mood constancy	91
MATTHEWS, GARETH B.	On not being said to do two things	204
MAYBERRY, THOMAS C.	The *same* psychological state	122
MURPHY, JEFFRIE G.	Three mistakes about retributivism	166
O'SHAUGHNESSY, BRIAN	The powerlessness of dispositions	1

P433

Place, U. T.	The infallibility of our knowledge of our own beliefs	197
Puccetti, Roland	Mr. Brennan on persons' brains	30
Radford, Colin	Characterizing-judgements and their causal counterparts	65
Resnick, Lawrence	Evidence, utility and God	87
Robinson, G. S.	Popper's verisimilitude	194
Robinson, Richard	Begging the question, 1971	113
Smart, Brian	Can disembodied persons be spatially located?	133
Squires, Roger	Are dispositions lost causes?	15
Taylor, Gabriele *and* Wolfram, Sybil	Virtues and passions	76
Thomas, William J.	On behalf of the Skolemite	177
Van Evra, James	On death as a limit	170
Whiteley, C. H.	Empirical, analytic, evaluative	118
Wilkins, Burleigh T.	The limits of historical knowledge	58
Wilkins, Burleigh T.	Concerning 'motive' and 'intention'	139
Wolfram, Sybil	(*See* Taylor, Gabriele)	76
Zemach, Eddy M.	Time and time' again	62

ANALYSIS 31.2 DECEMBER 1970

105 13
AN
v.31, no.2-6
Dec 70 - June '71

WITTGENSTEIN, PHILOSOPHY AND LOGIC

By Ilham Dilman

I

WITTGENSTEIN'S chief interest was always in logic and in mathematics, and he considered the study of logic to be fundamental to philosophy. His discussions in *Philosophical Investigations*, for instance, always gravitate towards questions in the philosophy of logic. Unless one sees how much these are at the centre of his interest one will miss much of what he says on other issues.

By a 'study of logic' I do not mean a study concerned with casting propositions into symbolic form, mapping out their implications, formalizing arguments, developing methods for detecting fallacies in them and checking their validity, discovering logical proofs, constructing formal systems. On the contrary, Wittgenstein repeatedly warned against the tendency towards formalization in philosophy: 'A piece of mathematics cannot solve a problem of the sort that trouble *us*' (*Remarks on the Foundations of Mathematics* V, sec. 19). The kind of study of logic to which Wittgenstein devoted himself is one that aims at understanding questions such as the following: What makes it possible to construct a formal system such as, for instance, Russell and Whitehead constructed in *Principia Mathematica*? (see *Tractatus* 5.555) What kind of relation is there between the principles of logic and the propositions that occur in our arguments? How is it that nothing actual can contradict the propositions of logic? What gives them the kind of generality they have? How are we to understand the kind of sway they have on our thoughts? An investigation into questions of this sort is what I mean when I speak of a *philosophical* study of logic.

The difference, however, between a philosophical and a non-philosophical study of logic is not one of method and, of course, Russell too in *Principia Mathematica* was concerned with the nature of logic, although his work looked very much like a treatise in logic and mathematics. Let us suppose, for instance, that one is concerned to construct a purely formal system, starting with a few definitions by means of which certain signs are introduced, and a few axioms. At one point one will be concerned, for instance, with whether one's system is complete, whether it involves any inconsistency. If one discovers any inconsistency, any hidden contradictions, one will try to make appropriate alterations or additions with a view to removing the contradictions. Here one is employing logico-mathematical techniques and in that sense one could be said to be doing logic. So far no philosophy has come into the picture. Yet in *Principia Mathematica* Russell and Whitehead are certainly doing something similar: trying to formalize and to unify the whole of mathematical thought, to derive it from a few non-mathematical axioms, to prove that every part of it does follow logically from these axioms, and to remove any contradictions that may arise in it. In this respect they may be said to be doing logic and mathematics, to be engaged in a logico-mathematical investigation. For surely the question whether or not the propositions of mathematics follow from logical axioms, the question of how a contradiction may be resolved, are purely logico-mathematical questions.

Yet the motive-power of this whole venture comes from philosophy. It was, I think, largely with philosophical questions in mind that Russell and Whitehead embarked on this venture. Their aim was not to extend mathematics, nor to improve it for practical purposes. If they said that they were concerned to remove certain defects in mathematics, then one ought to remember that these were not defects which interfered with those using mathematics to solve problems. Again, if one wishes to say that they made mathematical discoveries one should remember how these differ from other mathematical discoveries. Russell himself states this difference lucidly in his *Introduction to Mathematical Philosophy*. In the opening paragraph of that book he writes:

> Mathematics is a study which may be pursued in either of two opposite directions. The more familiar direction is constructive, towards gradually increasing complexity: from integers to fractions, real numbers, complex numbers; from addition and multiplication to differentiation and integration, and on to higher mathematics. The other direction, which is less familiar, proceeds, by analysing, to greater and greater abstractness and logical simplicity; instead of asking what can be defined and deduced from what is assumed to begin with, we ask instead what more general ideas and principles can be found, in terms of which what was our starting-point can be defined or deduced. It is the fact of pursuing this opposite direction that characterizes mathematical philosophy as opposed to ordinary mathematics. But it should be understood that the distinction is one, not in subject matter, but in the state of mind of the investigator. The distinction between mathematics and mathematical philosophy is one which depends upon the interest inspiring the research, and upon the stage which the research has reached; not upon the propositions with which the research is concerned (pp. 1–2).

So mathematical philosophy, as Russell practised it, uses purely mathematical and logical techniques, but differs from ordinary mathematics in the direction in which it proceeds and in the interest which inspires the research. Characteristically Russell regarded this as part of his programme of making philosophy "scientific", "precise", and non-speculative. In the preface to his *Introduction to Mathematical Philosophy* he says: 'Much of what is set forth in the following chapters is not properly to be called "philosophy", though the matters concerned were included in philosophy so long as no satisfactory science of them existed. The nature of infinity and continuity, for example, belonged in former days to philosophy, but belongs now to mathematics.' He means that these questions can be treated in a purely formal way and are susceptible of final and definite answers. Wittgenstein's way of thinking about these *same* questions in both the *Investigations* and the *Remarks* were so different from Russell's, indeed so "unmathematical", that Russell did not recognize that Wittgenstein was continuing to be concerned with the questions they had discussed together when Wittgenstein was a young man.

In what sense were Russell's formal researches inspired by philosophical interest and directed to philosophical difficulties? In the last chapter of his *Introduction to Mathematical Philosophy* Russell says that many mathematicians have learnt a technique, which they employ with great skill, 'without troubling to inquire into its meaning or justifications' (p. 194). It is well known that such questions as 'What does it *mean* to say e.g. that one is in pain?', 'What kind of reality does a mental image have?', 'What is the ultimate *justification* of moral judgments or religious beliefs?', 'What are our reasons in the end for believing in the existence of the material things that confront us—the chair we are sitting in, the desk we are writing on?', 'What are we really saying when we say of someone that he feels angry?', 'What kind of claim are we making when we say that what someone did was a shabby thing to do?'—these are characteristically philosophical questions. Philosophers who asked questions like these have often wanted to look for *deductive connections* between one kind of statement and another—for instance, between statements about chairs and tables and reports about our sense impressions, propositions about feelings and desires and descriptions of human behaviour, moral judgments and statements of fact. It often seemed to them that unless such a connection could be established they would be forced to admit that we cannot have any binding reason for the kind of judgment we constantly make.

What drove Russell to base mathematics deductively on non-mathematical logical propositions is very similar to what so often drives philosophers to find deductive connections between one kind of proposition and another, to find a deductive justification for the principles that govern various forms of reasoning. There is an additional feature in the case of mathematics which equally threatens to give rise to sceptical doubts: mathematical propositions hang together in a very special way and are not independent of each other. Hence it seemed to Russell that unless one can provide a general guarantee that it is impossible to derive contradictory propositions from any one or more mathematical propositions one cannot be certain that one is ever calculating in using mathematics. His concern to prove that mathematical propositions form a consistent system and his concern to prove that mathematics is finally indistinguishable from logic, therefore, stem directly from his concern to meet philosophical scepticism. Unless mathematics is finally indistinguishable from logic, it seemed to Russell, one would have to admit that it is *arbitrary*. This, it seemed, would open the door to conventionalism with regard to mathematics and finally to scepticism. Unless one can guarantee that the different parts of mathematics will under no circumstances come into conflict with each other, Russell thought, there is no denying that they may be just arbitrary techniques. Thus the attempt to found mathematics on logic and the attempt to prove it consistent are part of the same concern.

II

Speaking of the indistinguishability of logic and mathematics in *Introduction to Mathematical Philosophy* Russell says:

> The proof of their identity is, of course, a matter of detail: starting with premises which would be universally admitted to belong to logic, and arriving by deduction at results which as obviously belong to mathematics, we find that there is no point at which a sharp line can be drawn, with logic to the left and mathematics to the right. If there are still those who do not admit the identity of logic and mathematics, we may challenge them to indicate at what point, in the successive definitions and deductions of *Principia Mathematica*, they consider that logic ends and mathematics begins. It will then be obvious that any answer must be quite arbitrary (pp. 194–5).

This is a typical case where it is claimed that a philosophical conclusion has been proved by demonstrative arguments. It is impressive, it seems incontrovertible: it seems that here we have a philosophical conclusion which has really been proved demonstratively. So much so that it seems almost as if the conclusion proved was all along a mathematical one. Yet I would say, as Wisdom said of Moore's proof, that although it is logically impeccable it is nevertheless sophistical.[1]

One can contest the philosophical conclusion that Russell argued for without contesting the mathematical proof he gave. One may ask: Is the reasoning we employ in solving arithmetical problems, for instance, *the same* as formal, deductive reasoning which involves no mathematical ideas, *if* it is established that the principles that govern the one and those that govern the other are deductively connected? Does deductive connection give us an adequate criterion of identity? This is a question which Russell does not raise and to which he assumes the answer would be obviously in the affirmative. But this unspoken assumption can and must be questioned, and it cannot be investigated by the kind of deductive procedure employed by Russell. In other words, there is an important question about what it is that Russell's and Whitehead's proof establishes: Granted that one formula is shown to follow from another, what does this actually prove? That these formulae follow from those ones is perhaps unquestionable. But is Russell's conclusion about the identity of mathematics and logic itself unquestionable? I do not think so.

---

[1] Wisdom, 'Moore's Technique', *Philosophy and Psycho-Analysis*, pp. 130–1.

Put it this way: Perhaps Russell does succeed in showing that all mathematical procedures are in accordance with the rules of logic. All this means is that 'mathematics is a logical method' (*Tractatus* 6.2, 6.234). As Wittgenstein puts it in the *Remarks*: 'So much is true about saying that mathematics is logic: its movement is within the rules of our language' (I sec. 164). Does this mean that mathematical and non-mathematical deductive reasoning are identical? The answer to this question is by no means obvious. As Rhees once put it: The questions (i) whether you can show that all mathematical procedures are in accordance with the rules of logic, and (ii) whether the mathematics that we learn and use, all of it, is something which could have been developed from the principles of logic—these are quite different questions. Wittgenstein would have said that the identity of a form of reasoning cannot be decided in separation from the kind of context in which it is exercised and the kind of activities into which it enters: it cannot be decided on purely formal grounds.

In any case, need mathematics have the kind of foundations on which Russell tried to base it? Without such foundations are mathematical techniques arbitrary? Does their employment in a particular case lack the force of calculation? What gives their employment the force of calculation? What is the difference between carrying out a calculation and playing a game? Without questioning the steps in the formal proofs which Russell and Whitehead developed, Wittgenstein questions their philosophical conclusions and the unspoken assumptions behind their way of reaching them. More radically he questions Russell's whole conception of logic and mathematics of which the philosophical conclusions in *Principia Mathematica* are a part.

III

Can there be a deductive guarantee that the methods developed by mathematicians will not be upset by the appearance of some case for which they have not provided? To Russell it seemed that unless such a guarantee can be provided, unless it can be proved that what was for him the foundations of mathematics can *under no circumstances* lead to a contradiction, one has no right to trust mathematics. In *My Philosophical Development* he writes of how in 1901 he discovered a contradiction in mathematics. He explains there that it was a consideration of Cantor's proof that there is no greatest cardinal number that led him to this discovery.

Some classes, Russell explains, are not members of themselves, while some others are. For instance, the class of the books in this room (the example is his) is not another book in this room, but the class of things that are not books is itself one of the things that is not a book. Now take, says Russell, the class of classes that are not members of themselves. Is this class a member of itself? (A) If it is, then like the rest of its members, like the class of books, for instance, it must be a class that is not a member of itself. (B) If, on the other hand, it is not a member of itself, then it must lack the defining property of the class, and so it must be a member of itself. There are only these two alternatives and each leads to its opposite. In other words, the class of classes that are not members of themselves can neither be a member of itself, nor not be a member of itself, and yet it must be one or the other. In short, it cannot be what it must be. This is paradoxical. The concept of a class of classes that are not members of themselves hides a contradiction. Yet here is a concept which is at the centre of Russell's account of number and therefore, according to Russell, one which occurs at the ground level propositions which support the whole superstructure of mathematics.

Russell writes: 'I wrote to Frege about it, who replied that arithmetic was tottering. He was so disturbed by this that he gave up the attempt to deduce arithmetic from logic, to which, until then, his life had been mainly devoted. For my part, I felt that the trouble lay in logic rather than in mathematics and that it was logic which would have to be reformed.' Thus the theory of logical types (writes Russell) 'recommended itself to us in the first instance by its ability to solve certain contradictions'.[1]

There is one point on which Russell and Frege are in agreement, namely that the contradiction must be removed. Frege said that arithmetic was tottering; Russell thought that contradictions, if not removed, would vitiate the whole of human reason. He thought that either they can be removed or universal scepticism is inescapable. In his later work Wittgenstein questions not only the formulation of this paradox (*Remarks* V, sec. 29) and the connection between the notions of class and number,[2] but the whole conception of logic and reasoning behind this view. In the *Investigations* he writes:

> It is the business of philosophy, not to resolve a contradiction by means of a mathematical or logico-mathematical discovery, but to make it possible for us to get a clear view of the state of mathematics that troubles us: the state of affairs *before* the contradiction is resolved.

---

[1] Second Introduction to *Principia Mathematica*, Cambridge Paperback, p. 37.

[2] In the *Tractatus* he had already questioned Russell's conception of number (see especially *Introduction to Mathematical Philosophy*, pp. 132–3) which requires the existence of aggregates and the Axiom of Infinity.

> The fundamental fact here is that we lay down rules, a technique, for a game, and that then when we follow the rules, things do not turn out as we had assumed. That we are therefore as it were entangled in our own rules.
>
> This entanglement in our rules is what we want to understand.
>
> It throws light on our concept of *meaning* something. For in those cases things turn out otherwise than we had meant, foreseen. That is just what we say when, for example, a contradiction appears: 'I didn't mean it like that.'
>
> The civil status of a contradiction, or its status in civil life: there is the philosophical problem (sec. 125).

This is not only a remark about philosophy, but also one about logic.

The desire for a general guarantee that the rules or principles in accordance with which we proceed when we calculate will *in no circumstances* lead to a contradiction, the desire to cope with every conceivable situation in advance and not to cope with difficulties as they come up—this is part of what Wittgenstein called 'a craving for generality'. He thought that the attempt to systematize logic in abstraction from the practical situations in which it is employed, to look at logic in separation from the way the language we speak enters our lives in different situations, is misconceived. This way of looking at logic, exemplified by the mathematical logician, contains or encourages a misconception about the very nature of logic (see *Remarks* IV, sec. 48).

It is this misconception which is at the centre of Wittgenstein's attention when in the *Investigations* he discusses how a rule or formula determines the steps taken by someone who is observing the rule or using the formula. He argued there that rules and formulae, and also gestures and definitions, examples and illustrations, hints and pictures, can always be meant and so taken or understood in a variety of ways. Hence the statement of a rule, a definition, cannot *itself* give us a criterion of the right way to act, the correct way to use the word: 'Does the sign-post leave no doubt open about the way I have to go? Does it shew which direction I am to take when I have passed it? But where is it said which way I am to follow it? And if there were, not a single sign-post, but a chain of adjacent ones or of chalk marks on the ground—is there only *one* way of interpreting them?' (*Investigations* sec. 85) What determines which is the correct way of understanding a gesture, rule, or definition, is *general practice*. It is there that how the rule or definition is meant and so how it is to be taken is to be found. A particular person will be able to take part in the practice if he can be trained in certain ways—and that means if he naturally reacts or is brought to react as a matter of course to certain instructions, gestures and examples in ways which participants in the practice regard as being in agreement with their own reactions in similar situations. The important point for the moment is that it is the practice, founded on a consensus of reactions, that enables us to say of a person that he is following a particular rule, say, in arithmetic, that what he does at one step or on one occasion is consistent with what he does on another. It is this that enables us to distinguish between a correct and an incorrect procedure; such words as "correct" and "consistent", such notions as accordance and conflict, mean nothing apart from an established way of doing things: 'A person goes by a sign-post only insofar as there exists a regular use of sign-posts, a custom' (sec. 198). In short, what people confronted by a sign-post or formula *actually do in particular cases* determines what the sign-post says, what the formula means: 'Not only rules, but also examples are needed for establishing a practice. Our rules leave loop-holes open, and the practice has to speak for itself' (*On Certainty*, sec. 139).

Derivation in accordance with a formula is an abstract and mechanical procedure. But what are and what are not correct derivations—this is determined *in the end* by what people do in particular situations. The lines followed mechanically by a purely formal, abstract procedure cannot overreach actual practice. A proof, an abstract procedure, can lay down 'new tracks' along which the practice may be extended. But there is no compulsion on the practice to develop just this way, in just this direction. To put it differently, if people did not actually, in their lives, actions and speech, follow the lines suggested by such an abstract procedure there would be no grounds for saying that they were not being true to their commitments. It could not be said that they were being inconsistent with any part or feature of their practices in refusing to follow the guide lines suggested by this purely formal procedure. For, as we noted, the notions of consistency, necessity and commitment have sense only within an established way of doing things. The meaning which the formal logician gives to words and symbols which are used in certain ways by people in certain situations in drawing inferences which take him beyond actual use is *not* the only possible meaning that could be seen in or given to these words and symbols and the rules which govern their use. The people who use them can always tell the formal logician that they 'didn't mean them like that'.

So if by a purely deductive procedure a contradiction were derived from the rules in accordance with which we reason this need not show that there is anything wrong with these rules. As Wittgenstein puts it: 'The sign-post is in order—if, under normal circumstances, it fulfils its purpose' (*Investigations* sec. 87). The point is that you cannot talk of whether or not the sign-post is in order in separation from the circumstances in which it is used and what role it plays there. To elaborate an example of Wittgenstein (*Remarks* II, sec. 79): If someone asks you the way to the hospital, you may point to the right or to the left according to which way the hospital lies. But if you were to point both to the right and to the left he would not know which way to go. We would say that the directions you had given him were contradictory. But if the street were circular then these would be alternative directions and they would not contradict each other. Whether or not they are contradictory, whether or not we should avoid pointing both to the right and to the left in answer to the question 'Which is the way to the hospital?', depends on the circumstances.

This is equally true of the rules of logic. The formal logician needs to be reminded that 'it is for practical, not for theoretical purposes, that the disorder is avoided' (*Remarks* II, sec. 83). That is why Wittgenstein directs the philosopher's attention to 'the civil status of a contradiction.' Here we should not forget that the contradictions in question are not contradictions in an argument, but contradictions generated by the principles of logic. Such a contradiction can be 'sealed off' (II, sec. 80), if we actually come upon it, by specifying that the principles or rules in question are not meant to be taken in that way.

We have already seen that the meaning of any rule lies in the practice in which it is observed. If we think that we can and must 'seal off' every possible disorder or contradiction, we have once more given way to that 'craving for generality' which Wittgenstein was combatting. We should, therefore, ask ourselves what 'every possible disorder' could mean here? As Wittgenstein puts it very succintly:

> If I had to fear that something somehow might at some time be interpreted as the construction of a contradiction, then no proof can take this indefinite fear from me.
>
> The fence that I put round contradiction is not a superfence (*Remarks* II sec. 87). (Also see *On Certainty* secs. 26–7).

In the *Investigations* Wittgenstein wrote: 'That is not to say that we are in doubt because it is possible for us to *imagine* a doubt' (sec. 84). Similarly one could say: That is not to say that the mathematics we use or our logic is defective because it is possible to imagine a contradiction, *i.e.* to derive one from its rules in a purely abstract manner (*On Certainty* sec. 392). The possibility of such a contradiction is not anything we could remove or need to fear: 'It is not the eternal correctness of the calculus that is supposed to be assured, but only, so to speak, the temporal' (*Remarks* II sec. 84).

The problem, as Russell saw it, was *how to reform logic* so as to make it consistent, so as to remove the contradictions he came upon in the course of his work. This is a problem in mathematical logic (*Remarks* V sec. 16). What makes it of philosophical interest is that the impetus to reform logic comes from a certain conception of logic. It is *this* conception which Wittgenstein was concerned to change (*Remarks* II sec. 82). It is this conception which is behind Russell's philosophical problem—namely how scepticism is to be met. He thinks that it could be met by removing the contradiction. This, in turn, is a mathematical problem. Wittgenstein, on the other hand, suggests that if one took a different conception of logic, if one gave up the habit of 'looking at language without looking at the language-game', one would see that there is no need to worry as Russell did. In other words, the scepticism which Russell wanted to counter comes from a misconception of logic, from a 'misunderstanding of the logic of language' (*Investigations* sec. 93)—a misunderstanding concerning the relation between logic and language. So the scepticism in question will be met not by reforming logic, but by removing the misconception.

Removing such misconceptions means coming to a better understanding of the nature of logic. This is the sort of study that is at the centre of Wittgenstein's interest. Such a study, as we have seen, cannot be pursued in a purely formal and abstract manner. Equally important, I submit, is the fact that the questions that such a study investigates are those which a deep investigation of almost *any* philosophical question is bound to bring us up against. It is in this sense, I believe, that a study of logic is fundamental in philosophy. Hence Wittgenstein's interest in logic and mathematics was not just an interest that he happened to have. In it we have a measure of the depth of his interest in philosophy. In fact, Wittgenstein's interest in logic and mathematics and his interest in man and in human life and culture interpenetrate one another. When one sees the kind of interest he had in logic one will not find this surprising.

University of Hull

A MODEL FOR THE RELIGIOUS PHILOSOPHY OF D. Z. PHILLIPS

By ALAN BRUNTON

IN my philosophical journeyings I have always been most sorry, no doubt quite unnecessarily, for the unfortunate contributor to *The Philosophy of G. E. Moore* who, having written a most elaborate essay on the logical structure of Moore's Ethics, was informed in Moore's reply (p. 623) that his Ethics had (at least in the sense suggested by Professor Edel) no logical structure whatever.

I shall not make the same mistake with Mr. Phillips, though I shall risk presenting a model, by which I mean little more than a *simplified* logical skeleton or framework, by means of which the direction of his thought might be somewhat clarified. I wish, that is, to see the wood by ignoring the intricate traceries of the trees.

What model do I wish to suggest? Oddly enough, no doubt, at least at first blush, that which can best be described as Professor Ryle's adverbial model for the concept of 'Mind'. I call this a model, since to understand Ryle's book without its wealth of detail *is* to understand a radical change-over from 'mental' concepts understood substantivally—the Will, the Mind, the Imagination—to those same concepts understood adverbially—wilfully, intelligently, imaginatively.

Of course, I must insist that the analogy I wish to press between Ryle and Phillips is a restricted one. No doubt their subject-matter, styles and interests are very different. *All* I wish to suggest is that, if we interpret Mr. Phillips' statements about God adverbially, with the corresponding modifications for religious thought in general (the concepts of 'prayer' 'eternity' 'Judgment' 'Heaven' 'Hell' would all be affected), then we shall be logically illuminated.

What can be said in defence of the analogy I have suggested? First a few words as to what the analogy is. I am suggesting that, just as Ryle argues that there are not Minds and Bodies, but rather persons who act intelligently, stupidly, imaginatively, woodenly, Phillips argues that there are not persons *and* God (Himself either a Person or else a strangely, non-finite, non-three-dimensional Super-Person), but rather persons who have the divine spirit (know God), and those who don't.

Having the divine spirit, which admits of degrees, can be equated, at its best, with loving in a completely selfless *manner* (hence an adverbial account) as exemplified in the life of Jesus. Christians have this as an ideal which they, more or less, come up to at times, but which eternally exists for them as *the* way, *the* standard for their hearts and their heads.

If this is a worth-while analogy it should be possible to do more than state it. Two things, I think can be done. First, I should like to suggest that both Ryle and Phillips are naturally sensitive about the apparent reductionist tendencies of their arguments when misunderstood by the Philistine. It will be remembered that Ryle did not object to the characterisation of his book as behaviourist in tendency as long as this was not misconstrued as a kind of reductionism. Indeed, he made it clear that his whole enterprise, his plotting of the logical geography of our 'mental' concepts, was not meant as a revisionist exercise perpetrated on common-sense thought or expression. There's nothing wrong with 'He's got a good mind' if it doesn't mislead. Above all Ryle made it crystal-clear that he was not getting rid of things, least of all of Mind, Intelligence, Imagination, Wit, Will-power—but rather that he was liberating these rich concepts from their metaphysical, Cartesian strait-jacket, by giving them a new, adverbial, lease of life.

Similarly any suggestion that Mr. Phillips is a 'Death of God' man, or even that he is with the Demythologisers, would be stoutly resisted. His wish is to emphasize the richness of religious discourse, its home in a long tradition. 'Pictures', metaphors, parables, all play their part in the form of life which is Religion. There is no desire to strip the Poetry and leave only the Prose.

For Ryle, the babbling which the sound of my voice might convey to an Eskimo with no English does not have its meaning added as a separate ingredient. For Phillips a form of words does not become a prayer by *adding* God as the addressee. Yet meaningful discourse is not the same as babbling, and praying to God[1] is very different from a 'secular' desire for a lucky break and better times around the corner.

My second point is this. I have argued that neither Ryle nor Phillips are reductionists. They don't, in a sense, want to get rid of anything. Yet neither *The Concept of Mind* nor *The Concept of Prayer* were written as idle amusement. Some things had to go. On the one hand, the Cartesian Myth with its attendant ghost had to be exorcised. On the other, the Eschatological God, wiping away on the Last Day the tears of the just and repentant and augmenting those of the irretrievably wicked, had at least to be re-translated into adult terms.

In both cases our philosophers could argue that they were taking nothing real away. A logical child, with a little tuition and encouragement, begins to see conceptual difficulties in Santa Claus' simultaneous position in a million chimneys. Have we taken anything away? Certainly not Santa Claus. Not only does he not exist, but too many conceptual threads would have to be broken to suggest that he might.

Similarly, to take away a potentially disembodied soul or a Super-God, 'out of Space, out of Time'—is this to take anything away? Not, perhaps, if one has shown the logical ineptitudes involved in their postulation. [Yet one may remark on this point that Mr. Phillips has more difficulties to contend with than Professor Ryle. *The Concept of Prayer* unlike *The Concept of Mind* is not a predominantly secular book. If, for example, one is to write off the after-death hopes of Christianity's martyred saints and tormented sinners as superstition, one *may* be obliged by clear thinking to do so. Yet, if the beliefs, however illogical, of millions of Christians, are the gnat at which his critics are straining, then Mr. Phillips must at least admit that it is a fairly camel-sized one.]

---

[1] 'Praying to God' is like 'He has a good mind', unobjectionable when not misconstrued.

Is the analogy I have suggested a fruitful one? At least I think it will serve two useful purposes.

Firstly, a bold acceptance by Mr. Phillips of the adverbial model would put paid, once and for all to what, for him, must by now be an exasperating type of criticism and misunderstanding of his position. Stock criticisms of *The Concept of Prayer* (Professor Flew's one-sixth part of his omnibus book review in *Mind* for April 1967 will do as an example) complain, partly of obscurantism, partly of the wide gap between Phillips' view and what is actually believed by the Saints, Church Fathers, and professing Christians in general.

Now the obscurantism amounts to little more than the fatiguing effects on the reader of Phillips' continuing to talk about God, a quality, better perhaps a whole dimension, of loving-kindness, in the traditional terms in which God figures as a substantive with a capital letter.

Professor Flew finds it hard to believe (p. 295) what 'Phillips seems constantly to suggest, even if he never actually says, that there is no room for questions about the truth of religion as opposed to questions about what is truly religious'. Professor Flew's diagnosis is, in this point, correct and his belief should be more firm. D. Z. Phillips means exactly this; and, to be fair, his writings imply, in a pretty strong sense, that this *is* what he means.

As for the other problem, the correspondence or lack of it between Phillip's concept of the Christian Religion and the views of its professors; at least, if we accept the adverbial analysis we shall see more clearly where we stand. Here again Phillips is not disingenuous. In many places in his writings he says boldly, in effect, that he rejects the notion of personal survival with its attendant possibilities of post-mortem agony or bliss. He also makes it clear that he rejects *that* concept of God the Father, which makes it anything remotely like the name of a logically identifiable *individual* of however unique a character. That such an Individual should be postulated as infinite, or as the essence of Being rather than *a* being, would not, for Phillips add any intelligibility nor diminish in any way the offence against logical respectability.

An individual, he would say, or even Being, of however exalted and mysterious a nature, cannot logically fill the role of eternal love.

Now this is no doubt *unorthodox*, though one can always argue about the level on which the Christian faith has been appreciated. Even if, however, after we have grasped some of Mr. Phillips' Kierkegaardian insights into the nature of the Eternal, we feel that his views are still unorthodox; we shall, with the adverbial analysis, at least know where we are. If Jesus and the more religiously profound of Christians have understood that eternal life is a dimension of living now and not a promise for after death; if they have understood that God's love is a manner of loving and not a Super-Individual's care for our eventual welfare, then this is what they have understood. Whether this is a *plausible* view will depend *both* on what Jesus and many other individuals have said *and* on our interpretation of their words. Here one, including Mr. Phillips, can only be honest and remember to distinguish between what we take to be the beliefs of individuals and what we regard as the most profound elements in those beliefs and the best way of interpreting them.

The first purpose of my analogy then has been to make clear, at least in bold outline, what Mr. Phillips is saying; so that at least we know, in rough, what we are attacking, if this is what we want to do.

Secondly, I regard a simplified view of the structure of Phillips' thought as possibly clarificatory of his own position and the difficulties *he* finds in it.

In an article which is not meant to be exhaustive of the relevant issues, one example will suffice. In 'From World to God' (*Proceedings of Aristotelian Society* Supplementary Volume, XLI, p. 150) Phillips puts forward as one of his own difficulties the possibility of eternal love suffering defeat. This presents a logical difficulty since the eternal is not subject to a temporal, contingent factor. Can, then, affliction destroy the possibility of loving God?

A certain subtlety is required here. Phillips makes it clear that he is *not* discussing the harms which can, in any ordinary sense, *befall* a loving nature. To argue, as, for example, Winch does, that a good man cannot be harmed (*P.A.S.* 1965–66), is not to say that Jesus found crucifixion *enjoyable*. No! The point at issue is, whether a nature filled with the love of God[1] (loving in a completely unselfish spirit) can ever lose this gift of loving through intolerable adversity. It would seem that the answer should be 'No'. After all, anyone who regards adversity as a diabolical liberty perpetrated on his dear self, has not learned about God, and God's love. Since love expects nothing of itself, it can hardly be destroyed by injuries directed towards itself. Nevertheless, both Winch and Phillips are too honest to rest content with this answer. There *are* cases where an utterly loving nature can be so tortured by devilish wiles as to lose the power of loving. [One can imagine psychological pressures, or, perhaps better, tampering through drugs and surgical skills with a possible physiological basis or correlate of the personality.]

---

[1] It is important here to remember that we are talking of 'eternal' love, a love which does not provide its own restrictions by the limited nature of its interests and objectives,

Phillips' comment on this (though, to be fair, he is here only suggesting difficulties, not solving them) seems to cry out for that clear perception of the logic of his position which I am advocating. He writes 'Would believers say that God had deserted a man if the affliction which destroys his faith is itself the consequence of loving God?'

Well, to translate for the moment crudely into my terms, the question, without the capital 'G's and all that they entail, becomes this. 'Would believers say that a man who loves unselfishly has ceased to love unselfishly if whatever destroyed his love was itself the consequence of this love?' There is only one answer. 'Yes.' Would one say that a man who drives racing-cars superbly has ceased to be a driver if whatever resulted in his incapacitating smash-up was itself the consequence of his being on the racing-track? 'Yes', again.

A distinction must at least be made between the concept of 'love' and its exemplification, the concept of 'racing-car driver' and *its* exemplification, if we are to get clarity in this area. Whether, and in what sense, the *concepts* are related to Time, and are subject, not to change (that wouldn't make sense, in the case of 'eternal' love) but to annihilation or death by inanition, this is a fascinating question of a transcendental (Kantian) or limiting (Wittgensteinian) nature.

In my last comments I have been somewhat crude and provocative for two reasons. Firstly, I am only concerned in this paper with a simple, central point. Secondly, I should like to bring out into the open, in these final remarks, the possible dangers for my own case of over-simplification and a sacrifice of subtlety.

Have I, myself, been guilty of that very Reductionism which I have suggested both Ryle and Phillips wish to avoid? This would be the wrong *kind* of simplification, since any religious philosophy that has profited from the works of Sōren Kierkegaard could be nothing if not subtle.

I end by some kind of self-defence on this issue. Firstly, I don't intend to apologize for my insistence on *some* kind of 'adverbial' analysis. I do not have space to argue it here, but I would regard Mr. Phillips' position as *too* remote from that, say, of Professor Geach, for him to *avoid* an analysis of this kind.

Have I taken anything away from Mr. Phillips? I doubt whether I have, since he is certainly not a clinger-on to Santa Clauses.

(a) I don't get rid of his insistence that he is not just developing a religious philosophy of Harean 'bliks' or Braithwaitian attitudes. A person's love can fall within a tradition and have roots. It can be dependent for its nourishment on History, Family, and a host of other factors. There is no reductionism here.

(b) I don't reduce Religion to the sort of 'Utilitarian' morality which Mr. Phillips says it isn't, since a morality of selfless love is different from what he may call a 'stock' morality in exactly the way that he requires.

(c) Far from making the subtlety of Mr. Phillips' position suffer from my 'reductionism', I should hope on some occasion to show in greater detail that what remains of the transcendent, the paradoxical, the numinous, the ecstatic, the humble, by his principles, is exactly proportionate to the moral quality of the good man, the moral ideals of the average sinner, and the moral rejections of the wicked.

Yet this might still seem to be reductionism. Mr. Phillips might, at the end of all this, say, 'But, I never meant to say that Religion is *just* Morality. Indeed I have said the opposite.'

One can, indeed, appreciate Mr. Phillips' desire not to reduce the richness and subtlety of the religious life to what may be called 'stock' Morality. Does not Kierkegaard distinguish between the Aesthetic, the Moral and the Religious ways of life?

Indeed one has some sympathy with Professor C. K. Grant who, in his part of the Symposium 'From World to God' (p. 160), complained that selfless morality should not necessarily be equated with religious morality. No doubt Mr. Phillips would argue that Professor Grant's selfless nurse or doctor (p. 161) had a religious morality whether they knew it or not. Yet, like Professor Grant, one would have a slight repugnance to calling someone religious against his will simply because he did not (p. 160) ' "have expectations" as if he were a suitor in a Jane Austen novel'.

Yet, for all this, one sees what Mr. Phillips is getting at when he describes religious morality as being, as it were, a different dimension. Selflessness *is* paradoxical, mysterious. (How easily the logicians make their move: 'But what *sort* of a justification of selflessness are you asking for? A *prudential*[1] one?')

Yet, even here, I hope, the usefulness of my 'model' stands. It is the *ways* in which one is selfless, the context in which one sees the selflessness of others (saints rarely consider their own) that makes the difference. There are perhaps no extra items, no extra arguments to distinguish, say, Professor Prichard, who thought that unselfishness needed and could have no justification beyond itself, and Kierkegaard or Simone Weil, who, in one sense, thought the same. But, as Professor Wisdom might have said, 'The tone of voice! How different!'

---

[1] It is a pity that the word 'Prudence' conjures up a cross between an insurance company and a Round-head maiden. Perhaps, in a profound sense of the word, even the "saint" is prudent.

University College, Cardiff

FEATURE UNIVERSALS AND SORTAL UNIVERSALS

By MICHAEL DURRANT

IN his book *Individuals*, pp. 202 ff, Professor Strawson introduces us to what he calls 'feature universals' in contrast to his previously introduced categories of sortal and characterising universals. Examples of feature universals are *snow*, *water*, *coal* and *gold*: examples of sortal universals are *dog*, *terrier*, *animal* (p. 169): examples of characterising universals are *wise*, *warm*, *fights*, *talks*, *cries* (p. 169). A sortal universal supplies a principle for distinguishing and counting individual particulars which it collects: it presupposes no antecedent principle or method of individuating the particulars which it collects. Characterising universals on the other hand, whilst they supply principles of grouping, even counting, particulars, supply such principles only for particulars already distinguished or distinguishable in accordance with some antecedent principle or method. A feature universal is to be distinguished from both sortal and characterising universals: from sortal universals since no feature universal of itself provides a principle for distinguishing, enumerating and reidentifying particulars of a sort, even though they can be modified so as to yield several such principles: from characterising universals since no feature universal signifies properties or characteristics of particulars.

Now difficulties can certainly be raised concerning Strawson's criteria (p. 168) for deciding whether any given universal is or is not a sortal universal or is or is not a characterising universal; but this is not my interest here. I shall concern myself with the thesis which I understand Strawson to be advocating concerning the relation between feature universals and sortal universals—in that he advocates that it is perfectly conceivable that we should operate with a 'feature placing' language instead of a 'particular introducing' language in order to have a language in which singular empirical statements could be made. The possibility of such a conception is dependent upon showing that it is possible to introduce the idea of a feature universal where we now introduce the idea of a sortal universal, such that it would be possible to 'envisage a situation in which, instead of operating with the notion of the sortal universal, *cat* or *apple*, and hence with the notion of particular cats or apples, we operate with the notion of a corresponding feature and of placing' (p. 205). For example 'cat (here, there . . .)': 'more cat (here, there . . .)': 'cat again (here, there . . .)' instead of 'this cat': 'another cat': 'the same cat again' respectively (*vide* p. 207). On Strawson's view it is clearly not a necessary condition of introducing feature universals that one should introduce sortal universals, such that sortal universals or their introduction are logically prior to the introduction of feature universals. I shall contend that sortal universals *are* logically prior to feature universals and that Strawson in one passage at least has to admit this priority. I shall argue that insufficient grounds have been offered for his contention that it is possible to operate with a feature placing language in lieu of a particular introducing language.

Strawson realises that there are difficulties in his contention, but does not regard them as decisive. For example, he voices the objection (p. 205) 'Is it not perhaps the essential difference between, say, cats and snow, that there *could* be no concept of the 'cat feature' such as the theory seems to require, that any general idea of cat *must* be the idea of *a* cat *i.e.*, must already involve criteria of distinction and reidentification for cats as particular? These difficulties, though important are not decisive.' They are not decisive in Strawson's view for they do not show that it is logically absurd to suppose that there might be a level of thought at which we recognise the presence of cat or signs of the past or future presence of cat, yet do not think identifyingly of particular cats. Strawson's case that such a level of thought is possible is based on a false move: he wrongly infers from the consideration that one need not, in a given case, 'think identifyingly of particular cats', to the thesis that in such a case one can introduce that notion of 'cat' (feature universal) as opposed to 'a cat' (sortal universal). This move would be valid only if 'thinking identifyingly of particular cats' were a necessary condition for introducing the items '(a) cat', '(some) cats' into a proposition. But this is not necessary. One can be said to recognize the presence of a cat or cats without necessarily being able to say *which* cat or cats it is of which one recognizes the presence. I can well say 'Some cat has been here', *i.e.* I can genuinely introduce the item 'some cat' *without* being able to say *which* cat it is (without 'thinking identifyingly' of a particular cat) and further I can say 'some cat or other has been here', *i.e.*, I can genuinely introduce the item 'some cat' and yet the question 'which cat?' not sensibly arise. It does not make sense to raise the question 'which one?' in connection with *all* uses of 'some', as Geach has pointed out (*Reference and Generality* p. 9), and even in those cases where the question 'which one?' *does* make sense, it is not required that I should be able to specify which one in order to sensibly introduce 'some cat'; I simply may not know. In order to show that the thesis he is advocating here is not logically absurd in so far as this particular example goes, Strawson would have to show that there might be a level of thought at which we might introduce the feature universal 'cat', without having to introduce the notion of 'some cat' in either of the above senses: and this he does not do. Thus I contend that as yet he has not shown that there is no priority of sortal universals over feature universals.

Strawson does not rest his claim for the possible independence of feature universals merely on the above supposition; he has a positive argument (pp. 205–6). This argument depends on the coherence of the concept of the 'naming game'. The concept of the naming game is the linguistic counterpart to that level of thought in which one can be said, for example, to recognise the presence of cat without thinking identifyingly of particular cats. Strawson emphasises that all he is out to do here is to show that the idea of the 'naming game' is a coherent one, not that this 'game' is actually played. (This is designed to counteract in advance certain potential objections from followers of Wittgenstein: I shall not discuss the issue.) He says (p. 206) 'Playing the naming game may be compared with one of the earliest things which children do with language when they utter the general name for a kind of thing in the presence of a thing of that kind, saying "duck" when there is a duck, "ball" when there is a ball, *etc.*' The heart of Strawson's argument comes when he says: 'All that is required is the admission that the concept of the naming game is coherent, the admission that the ability to make identifying reference to such things as balls and ducks includes the ability to recognise the corresponding features, whereas it is logically possible that one should recognise the features without possessing the conceptual resources for identifying reference to the corresponding particulars.' His position, as I see it, is that granted that it is logically possible to recognise the features, without possessing the conceptual resources for identifying reference to the corresponding particulars, then feature universals can be regarded as independent of sortal universals. It is my contention that this argument fails for the following reasons:

(1) It implicitly depends upon the false move which I discussed above.
(2) It relies on an illegitimate switch from 'corresponding features' to 'features'.

It is noticeable that Strawson switches from 'corresponding features' to 'features' and this switch is important, since by 'corresponding features' would be meant *e.g.*, 'featheredness of ducks', or 'sphericity of balls', but by 'features' would be meant *e.g.*, 'featheredness' and 'sphericity'. One may agree that the ability to make identifying references to such things as ducks or balls includes the ability to recognise the corresponding features, *viz*. the featheredness of ducks and the sphericity of balls respectively *and* that it is possible to recognise featheredness and sphericity which is not *e.g.* the featheredness of ducks or the sphericity of balls, but in order for Strawson's case to hold it must be possible for one to recognise *e.g.* the featheredness of ducks and the sphericity of balls, simply from possessing the concepts of featheredness and sphericity. This however is not so possible, for the concepts 'featheredness of X', 'sphericity of Y', are not analysable into 'featheredness (of)' and 'X' and 'sphericity (of)' and 'Y'—since 'featheredness' and 'sphericity' are necessarily the featheredness and the sphericity of something or other and the identity of the featheredness or sphericity in any putative case is determined by the identity of what it is the featheredness or sphericity of. Rather, the phrases 'the featheredness of X' and 'the sphericity of Y' should be taken as analogous to 'the square root of X'—which certainly is not analysable as 'the square root (of)' and 'X'. (cf. Geach: 'Aquinas' pp. 77–8).[1]

(3) It is *not* possible to recognise the *corresponding* features here without being able to make the appropriate identifying references, for otherwise one would not be able to identify the cases of featheredness and sphericity under investigation as the featheredness of ducks and the sphericity of balls respectively and, if one could not make such an identification, then one could not be said to recognise. Yet in order to get coherence for his 'naming game' this is precisely the possibility Strawson requires. Strawson's case has only seemed possible here on account of the apparently innocuous transfer from 'corresponding features' to 'features': a transfer which has been anything but innocuous. With the failure of this argument comes the failure to produce an adequate case for saying that it is possible to introduce 'feature universals' without introducing 'sortal universals'. Indeed it is my contention, that feature universals of the type Strawson wants to introduce in this passage require the prior introduction of sortal universals for their identification.

(4) Finally Strawson seems to destroy his position out of his own mouth for, to the objection that there could be no idea of a cat-feature which would be distinct from, yet yield a basis for the sortal universal, *cat*, he replies: 'what this argument shows however, is not that the required type of general concept of cat is impossible, but rather that the concept must already include in itself the basis for the criteria of distinctness which we *apply to particular cats*.' Here, in order to explain how the required type of general concept of cat is possible, *i.e.*, how the feature universal 'cat' is possible, he has to invoke the idea of 'application to particular cats', that is, he has to introduce the sortal universal 'cat' in order sensibly to introduce his new concept of 'cat': as this is required, it cannot be claimed that the new concept of 'cat' ('cat feature') which he is introducing can "yield a basis for" the sortal universal 'cat' or be "distinct from" such a sortal universal. I thus maintain that the programme for replacing a particular introducing language by a feature-placing language is a non-starter as far as Professor Strawson's case is concerned.

---

[1] In *Three Philosophers*, G. E. M. Anscombe and P. T. Geach, Blackwell 1961.

University College, Cardiff

'EXISTS' AS A PREDICATE: A RECONSIDERATION

By James Child *and* Fred I. Goldberg

SINCE the criticisms of the ontological argument by Hume and Kant, it has been an almost unchallenged assumption that 'exists' is not a predicate. A number of years ago, however, Professors Nakhnikian and Salmon[1] attempted to show that one important argument against treating 'exists' as a predicate is not well-founded. The purpose of this paper is twofold. First, to consider and refute two other important, perhaps more widely held, arguments against treating 'exists' as a predicate and, second, to show a way in which such a predicate might be used to shed some philosophic light.

Nakhnikian and Salmon were most concerned to show that an argument given in slightly different forms by Broad, Ayer, and Wisdom is not sound. They claimed that when 'exists' is treated as a predicate, all affirmative existential propositions become tautologies and all negative existential propositions become self-contradictions. Nakhnikian and Salmon were able to show that the argument depends upon a mistranslation of the affirmative existential proposition and a misinterpretation of the negative existential proposition and that both existential propositions are in fact synthetic.

Nakhnikian and Salmon realize they must avoid the drastic consequences of importing non-existing entities into the universe of discourse. For this reason the existence predicate must be treated as universal, non-descriptive and redundant, *i.e.*, '$(\exists x)Hx$' implies '$(\exists x)(Hx \,.\, Ex)$' and is implied by it. Since 'E' is universal,

(1) $(x)Ex$

is true. The authors suggest that (1) be taken as a meaning postulate for 'E'. This extensionally defines the predicate and is all that need be said about its meaning within a specified language. It simply says that there are no non-existent entities.

I

The arguments of Nakhnikian and Salmon were constructed with reference to a logic without individual constants. Nonetheless, their claim that 'exists' can be treated as a predicate should hold in a logic with individual constants, since the latter can always be constructed from the former. However, such an alteration creates problems not foreseen by Nakhnikian and Salmon, for it appears to be inconsistent with Russell's theory of names and descriptions. Russell believed that

---

[1] George Nakhnikian and Wesley Salmon, ' "Exists" as a Predicate', *Philosophical Review* Vol. LXVI, No. 4, October, 1957.

P438

it was impossible to predicate meaningfully non-existence of a named entity. As he argued,

> It is only of descriptions—definite or indefinite—that existence can be significantly asserted; for, if 'a' is a name, it *must* name something; what does not name anything is not a name and, therefore, if intended to be a name is a symbol devoid of meaning . .[1] .

Thus, if '—Ea' is meaningless, 'E' violates the rule of significant denial and is not an acceptable predicate.

This argument may be interpreted in two ways. Upon first examination it appears to turn on a simple confusion. The claim that '—Ea' is not well-formed is based on the argument that '—Ea' says that *a* does not exist, and hence 'a' is not a legitimate individual constant. But if '—Ea' is not well-formed, it says nothing. It cannot be the case both that '—Ea' is not well-formed and that it makes an existence assertion.

It is not surprising that the argument understood this way is not sound for, if it were, the logically true formula '$(\exists x)\ (x = a)$' would as well be meaningless. This is so since its negation asserts that there is no entity which is identical with *a*, from which it follows that 'a' fails to denote. In fact, '$(\exists x)\ (x = a)$' and 'Ea' say exactly the same thing, viz., that *a* exists. For this reason both are logically true, and their negations are self-contradictory. The peculiarity of the predicate 'E' lies in the fact that, in a specified universe of discourse, it applies to everything.

The problem is that as an individual constant or proper name 'a' must refer to an extra-linguistic entity. When we remember that the extension of 'E' contains all entities over which the system is interpreted, '—Ea' says that the individual denoted by 'a' is not a member of the class of everything there is. Since 'a' must always denote something, '—Ea' is never true.

The source of difficulty appears to be a blurring of the distinction between truth and meaning. The state of affairs in which the individual denoted by 'a' is not included among the things there are follows not from the meaning of '—Ea', but from its *truth*.

This suggests what is perhaps a fairer interpretation of Russell's argument. He might not at all have confused meaning and truth as suggested above, but rather assumed the meaningfulness of 'Ea' to arrive by *reductio* at the incoherence of '—Ea'. If some negative existential propositions are true and 'Ea' is a legitimate expression, then '—Ea' is true for some *a*s. But certainly '—Ea' cannot be true, and thus if some negative existential propositions are true, 'Ea' is not legitimate.

---

[1] Bertrand Russell, *Introduction to Mathematical Philosophy*, (London, 1919), p. 179. A slightly different form of the argument appears in Russell's 'The Philosophy of Logical Atomism', *Logic and Knowledge* (London, 1956), p. 242–243. For an example of how this argument has been used in the discussion of existence, see William Kneale, 'Is Existence a Predicate?' *Readings in Philosophical Analysis*, edited by H. Feigl and W. Sellars (New York, 1949), pp. 36–37.

This argument, though stronger than the first version, is not without flaw. It does not follow from the fact that some negative existential propositions are true that some propositions of the form '—Ea' must be true. When one wants to assert non-existence of an object without contradiction, he must not name it, but describe it. This is, to say the least, consistent with Russell's own views about names and descriptions. 'The present king of France exists' is false and, since 'the present king of France' cannot be a name, it can only be formulated as a definite description.

The above analysis, it should be pointed out, is perfectly in accord with Russell's general attitude about existence.[1]

II

The most pervasive argument against the use of 'exists' as a predicate is also perhaps the most superficial. Both Ayer and Kneale claim that if one asserts that 'exists' is a predicate, one is committed to a reality beyond that of existence.

> The fundamental thesis of those who believe existence to be a predicate is that there is a sense of being logically prior to existence . . .[2]

The argument is roughly as follows: if existence is a property (and thus 'exists' a predicate) of an existent thing, then non-existence and other similar properties like fictitiousness must be properties of something as well. There must be something which does not exist in order for it to have the property of non-existence. So if existence is a predicate, it follows that non-existent beings '. . . have a mode of real being which is different from the mode of being of existent things'.[3]

Another way of formulating this objection is simply to assert that an existence predicate cannot be adopted in an ordinary extensional logic where the domain of discourse is limited to existing (actual) entities. The very possibility of the consistent use of the predicate as demonstrated above seems to vitiate this argument.[4]

However, the answer to the problem as posed by Ayer and Kneale is equally straightforward. When we take (1) to be a logical truth, it follows that there is only one kind of "being", and that any entity we can properly name has that "being", *i.e.*, simple extensional existence. We can, quite consistently, give definite descriptions of non-existent entities and even assert their existence. In doing so there is no difference between an extensional language with the existence predicate and one without it. It is the meaning postulate and the consequent contradictory character of '—Ea' that prevents the illicit question of the being of a non-existent entity from arising.

---

[1] In 'The Philosophy of Logical Atomism' Russell would have rejected 'E' because it can produce logical truths or contradictions in atomic statements. But this rejection is for ontological and even epistemological reasons and not for logical ones. 'E' is perfectly consistent with Russell's views in 'On Denoting' and *Principia Mathematica.*

[2] W. C. Kneale, *op. cit.*, p. 40.

[3] Alfred J. Ayer, *Language, Truth and Logic* (New York, 1946), p. 143.

[4] Though this argument suggests that 'exists' can only be a predicate in a modal logic, where possible but unactualized entities can be values of variables, the facts are rather different. Such a predicate is inconsistent with certain important modal logics, see Nicholas Rescher, 'On the Logic of Existence and Denotation', *Philosophical Review* (April, 1959).

III

Nakhnikian and Salmon felt that the universal and redundant character of the existence predicate made it rather uninteresting. We feel that, while essentially dispensable, the existence predicate can be used to make some interesting points.

Several of the most intractable philosophic problems have revolved around the existence of certain types of entities: universals, numbers, classes, theoretical entities of science, *etc.* Often these discussions take the form of suggested criteria of existence, that is, a specification of the necessary and sufficient conditions for a thing's existence. This issue is perhaps most prominent in contemporary philosophy in the discussion of the existence of theoretical entities. The question often turns on an examination of different criteria for physical existence. R. Harré[1] and Gustav Bergmann[2] formulate the problem in these terms, and Max Born[3] has attempted to specify a set of criteria not for the existence of all physical entities, but for the entities of microphysics. Probably the most widely discussed version of this approach is Ernest Nagel's.[4] In all of these cases the authors involved are suggesting one or more properties of things which would make them physically real, *i.e.*, define physical or microphysical existence. But this seems to make existence a property. Indeed, this objection has been made. In discussing the suggestions made by Nagel and Bergmann, Professor Maxwell says,

> One would hope that . . . over nine hundred years of debate and analysis have made it clear that existence is not a property. Now surely the characteristics of being mentioned in well-confirmed laws, being publicly perceptible, *etc.*, are properties of sorts; and if these comprised part of of the meaning of 'exist' then 'existence' would be a predicate (and existence a property).[5]

Maxwell considers this difficulty sufficient reason for abandoning the attempt to define existence in terms of other properties. But if 'exists' can correctly be treated as a predicate, it follows that the consideration of criteria of physical existence or criteria of microphysical existence is quite legitimate.

---

[1] R. Harré, *Theories and Things* (London, 1961).

[2] Gustav Bergmann, 'Physics and Ontology', *Philosophy of Science,* 28:1-4 (1961).

[3] Max Born, 'Physical Reality', *Physics in My Generation* (London, 1956).

[4] Ernest Nagel, *The Structure of Science* (New York, 1961).

[5] Grover Maxwell, 'The Ontological Status of Theoretical Entities', *Minnesota Studies in the Philosophy of Science*, editors, H. Feigl and G. Maxwell (Minneapolis, 1962), Vol. III.

Given a formal language and a domain of discourse over which to interpret it, the existence predicate is not descriptive within that language. But the predicate can have a metalinguistic interpretation which might be of interest. Such an interpretation gives meaning to the symbol by relating it to words in the metalanguage, which themselves may or may not appear in the object language. For example, Nagel suggests as a possible criterion of physical existence that an entity be publicly perceivable.[1] Recast with the explicit use of the existence predicate, this criterion would take the form of the following interpretation in the semantic metalanguage.

(2) 'E' applies to all and only publicly perceivable events.

Since 'E' applies to all of the entities within the domain of discourse of the object language, such an interpretation of the term would give the necessary and sufficient conditions for the existence of an entity within that particular domain of discourse and for that language. In other words, within a particular language the predicate is quite uninteresting, but when comparing various languages and domains of discourse, the notion of an existence predicate takes on some content.

It may be argued that in (2) the existence predicate is redundant since one could simply make the predicate 'perceivable' apply directly to all entities in the domain of discourse. But the adoption of the existence predicate does make explicit the point that *to exist* in this particular universe of discourse an entity must be perceivable. In addition, there are possible cases where the existence predicate would economize existence claims by conjoining or disjoining properties to make necessary and sufficient conditions for existence.

Needless to say, the existence predicate is not irreducible, and everything required can be said without its use and handled in the more traditional manner, asserting existence only with the existential quantifier. The important point is, however, that 'exists' can also be treated as a predicate; and for this reason one can talk in a standard extensional logic of criteria and conditions for and even definitions of 'existence', as Nagel *et al.* do, with a clear logical conscience.

---

[1] Nagel does not actually argue for the adoption of one or another of the alternative criteria he suggests. Rather he wants to show that since there are several possible criteria, any solution to the controversy is essentially arbitrary.

University of California, Santa Cruz
San José State College

THE LIMITS OF HISTORICAL KNOWLEDGE

By Burleigh T. Wilkins

ARTHUR C. DANTO'S *Analytical Philosophy of History* is an attempt to show (1) that historical writing is essentially narrative and (2) that an examination of the limits of historical narration will reveal the limits of historical knowledge. Although the model of "story-telling" is helpful in understanding some aspects of historical writing, the claim that historical writing is essentially narrative seems to me false. Probably the strongest claim which could be successfully defended here is that historical writing is frequently or often narrative in form. Moreover, I believe that even if (1) were true, Danto's examination of the limits of historical narration is misleading and could not help us discern the limits of historical knowledge; and it is this point which I shall try to establish in the present essay.

In the chapter 'Narrative Sentences' Danto invites us to try to imagine a case where we have a history of events written before the events about which it is written have happened. Suppose we come across a book titled *The Battle of Iwo Jima* which describes that conflict in '*minute detail*'. We then discover (a) that the book was written in 1815 and (b) that it is an entirely accurate account of the Battle of Iwo Jima.

> Everyone would say: this is a forgery. But if we found amongst Newton's papers a celestial map for the year 1960, and checked and found it wholly accurate, we would not suspect a hoax. We would not feel the unease which comes when a fundamental concept is threatened.[1]

Neither precognition nor prophetic visions will, according to Danto, explain the possibility of an account of the Battle of Iwo Jima having been written in 1815. Nor will an appeal to science help—today the historian lacks the sorts of theories and laws the astronomer has, and it is unlikely any such theories and laws were available in 1815. Moreover, the theory in question would have had to be enormously rich linguistically, as rich as ordinary language, for the theory would have been used to predict not just the occurrence of the Battle of Iwo Jima but that event 'under the enormously detailed description found in our controversial "history". There must be such sentences in the latter as "At 3:30 February 20, Sgt. Mallory, while arming a grenade, is killed by Pvt. Kito—the latter's fifth and only successful shot of that day".' (*Ibid.*, p. 178).

A history such as *The Battle of Iwo Jima* seems to preclude, Danto believes, the possibility of our being able to falsify predictions about what we will do and what will happen to us in the future, and this is why he finds his example so disturbing. According to Danto, 'It is a common enough thing to falsify predictions. . . The only way for the prediction to go through *is* for it to be discovered after the event'. If the book in question were discovered in 1944, the High Command might revise their plans, and Sgt. Mallory would 'see to it that he was elsewhere at 3:30 on 20 February'. The one book Danto claims we cannot imagine the men of Iwo Jima having in their hands is *The Battle of Iwo Jima*: 'Or rather we cannot both imagine their having it and the book remaining true' (pp. 179, 180).

---

[1] *Analytical Philosophy of History* (Cambridge, 1965), p. 173.

In my exposition of Danto I have noted that he contrasts history with astronomy, and it is here that I wish to begin my criticism. Undoubtedly, Danto is correct in saying that we would believe *The Battle of Iwo Jima* found in the papers of a nineteenth century author to be a forgery while we would not believe a celestial map for the year 1960 found among Newton's papers to be a hoax. Danto doesn't elaborate upon this point, but much of what he says later presupposes that an important difference between history and science has been brought out. But one might insist rather that at most an important difference between history and astronomy has been brought out, for other sciences are not as fortunate as astronomy which deals with a naturally isolated system and which perhaps in part as a result of this exhibits well-developed theories and laws yielding detailed predictions.[1] Even in the case of astronomy, however, one wonders what exactly in Newton's celestial map for the year 1960 would begin to approximate the sort of detail found in *The Battle of Iwo Jima*. Somewhat fancifully, one might suppose that the predicted collision of two stars might count as a prediction analogous to the warfare between Japan and the United States predicted by Danto's author; but what astronomical detail would stand comparison, *qua* detail, with the death of Sgt. Mallory at 3:30 on 20 February as a result of Pvt. Kito's fifth shot of the day? It may be that even astronomy does not exhibit that degree of detailed anticipation of future states of affairs which Danto thinks it noteworthy that history does not exhibit; but even if astronomy does, this is not characteristic of the natural sciences.

An exhaustive analysis of Danto's *Battle of Iwo Jima* example might proceed by taking the surprise that the discovery (and confirmation) of such a work would occasion and trying to locate this surprise on a Scale of Surprises. At the top of such a scale we might place the discovery of a married bachelor; next might come the discovery of a physical object which violates the law of gravity; near the bottom of our scale we might place such philosophically uninteresting items as the case of a man who lives to be a hundred and the case of a Maserati that goes a hundred thousand miles without a major overhaul. Then our task would be to attempt to explain why the married bachelor is more of a surprise than is the physical object which violates the law of gravity, and so on. Here, however, it seems sufficient to locate the *Battle of Iwo Jima* surprise below both of the above examples (no necessary truths or physical laws are violated by the discovery of *The Battle of Iwo Jima*; the most Danto would claim is that a "fundamental concept" is threatened) and yet well above the hundred year old man and the Maserati cases. In short, I want to grant Danto that the discovery of *The Battle of Iwo Jima* would be philosophically interesting. However, by sketching what I consider to be an equally bizarre surprise in the science of geology I shall attempt to show that in all philosophically relevant respects Danto's *Battle of Iwo Jima* surprise is of no particular concern to the philosophy of history (does not tell us anything distinctive about the limits of *historical* knowledge).

---

[1] Karl Popper writes: 'In general it is only by the use of artificial experimental isolation that we can predict physical events. (The solar system is an exceptional case—one of natural, not of artificial isolation; once its isolation is destroyed by the intrusion of a foreign body of sufficient size, all our forecasts are liable to break down.) We are very far from being able to predict, even in physics, the precise results of a concrete situation, such as a thunderstorm or a fire.' *The Poverty of Historicism* (Boston, 1957), p. 139.

Let us imagine the discovery of *The Santa Barbara Earthquake of 1925*, written in 1815, of course, in the papers of a nineteenth century geologist. This work gives an account of the earthquake which shook Santa Barbara, California, at 6:42 a.m. on June 29th, 1925, and of the thousand aftershocks which occurred in the next month; it provides also a comprehensive account of the effects of this quake and its aftershocks upon the natural environment in the Santa Barbara area. The details of *The Santa Barbara Earthquake* are as minute and as numerous as those in *The Battle of Iwo Jima*. For example, paralleling the story of Sgt. Mallory's demise, we are told that a given boulder (designated Boulder B-145 by our geologist) located on the northeast side of a certain slope (Slope S-19) of the Santa Ynez Mountains would, as a result of the twenty-third aftershock, come crashing down this slope crushing everything in its path; this detail, like all the details contained in *The Santa Barbara Earthquake*, is confirmed after the earthquake and its aftershocks occur.

I think that the discovery of such a work after the summer of 1925 would have occasioned as much surprise as the discovery of *The Battle of Iwo Jima* after the Battle of Iwo Jima, and that the explanation of the author's success would be equally baffling. Like Danto, we could rule out precognition or prophetic vision as *explanations* of our author's success; and, again like Danto in his treatment of the Iwo Jima example, we could argue that the appeal to science would not help either. Geology and all the sciences (with the possible exception of astronomy) today lack the theories and laws which would enable them to forecast in *minute detail* the results of what Popper calls 'concrete situations' as contrasted with the results of artificially isolated experimental situations; and there is no reason to suppose that geology was theoretically richer in 1815 than it is today.

If at this point we feel inclined to say that the surprises occasioned by the two discoveries of works written in 1815 are on a par on the Scale of Surprises, are there any disanalogies between the two cases which might still serve to show that the limits of historical knowledge are interestingly different from the limits of scientific knowledge? Perhaps what Danto says about the falsification of predictions about future states of affairs and how this is a common phenomenon where human affairs are concerned should be pursued further. However, while falsification may not be as common where natural phenomena are concerned, it isn't all that rare either. I should argue that, if it had been discovered in time and taken seriously, *The Santa Barbara Earthquake* could have been falsified to the same extent that *The Battle of Iwo Jima* is considered as falsifiable by Danto in his treatment of that work. For example, just as Sgt. Mallory might have tried, successfully, to be elsewhere at 3:30 on 20 February or the High Command might 'perhaps' have decided 'to alter their own plans' so an individual property owner or the County of Santa Barbara, concerned over the predictions in *The Santa Barbara Earthquake*, might have removed Boulder B-145 from Slope S-19 or else might have built a retaining wall around it.

Also, both *The Battle of Iwo Jima* and *The Santa Barbara Earthquake* could have been falsified by changes taking place among natural phenomena. In the case of the former work, an earthquake could have caused Iwo Jima to disappear beneath the Pacific prior to the Battle of Iwo Jima, thus eliminating Iwo Jima as a place over which two opposing armies might fight. In the Santa Barbara example, an earthquake prior to the quake of 1925 might have caused Boulder B-145 to disappear beneath the earth's surface.

The third possibility is that both *The Battle of Iwo Jima* and *The Santa Barbara Earthquake* might have been falsified by a mixture of changes occurring among natural phenomena and of changes initiated by human agency. In the Santa Barbara example, a prior earthquake might have caused Boulder B-145 to totter precariously above the San Marcos Pass Road, which could have led the County Commissioners to decide to have this boulder removed in the spring of 1925. In the Iwo Jima example a series of natural disasters might have led the Japanese to decide to withdraw from the island in 1944. I think, however, that the time has now come to surrender both *The Santa Barbara Earthquake* and *The Battle of Iwo Jima* to some deserving bibliophile for his Rare Books Collection.

After one final note; of course, there is an obvious difference between the falsification of predictions by human agency and by the operation of natural forces such as earthquakes. It may also be true that men are usually more able to falsify predictions about future human actions than about future natural occurrences. These differences, while important in other respects, are not, however, sufficient to establish Danto's point about the limits of historical knowledge which depends upon a contrast between what historical narration cannot and science allegedly can do. This contrast, I have argued, loses its effectiveness once we realize that the sciences, except perhaps astronomy, aren't able (as yet) to make minutely detailed predictions about concrete situations either.

University of California, Santa Barbara

TIME AND TIME′ AGAIN

By EDDY M. ZEMACH

IN an earlier paper ('Many Times'; ANALYSIS 28: pp. 145–151) I have said that the opponents of bitemporality (*e.g.*, Kant and Quinton) are partially right, since, given our way of remembering, it is necessary that we can acknowledge only one Time (even though we may recognize more than one Space). However, I argued, a conceptual framework is logically possible which could accommodate more than one Time. In such a framework there could be a spatial analogue to memory: the intelligent beings in this world (*e.g.*, Hurricane Betsie) would know non-inferentially (*i.e.*, "remember") some events (past, present and future) to the east of the remembering part, but not any events occurring west of the location of the remembering part. In such a world, an ordered series of events totally unconnected to the known series of events would not be explained as occurring in another Space, but rather as taking place in another Time.

Mr. Tobias Chapman ('A Note on Bitemporality'; ANALYSIS 30: 108–110), however, believes that such a bitemporal framework must be incoherent. The core of his argument, I think, is the following passage:

> HB (Hurricane Betsie) could enter and leave Time at the same moment though she would have a temporal part in between the leaving and the re-entry. But what about these temporal *parts*? Certainly her temporal part in Time′ is temporally connected to her previous temporal parts, it comes after them and could be measured by her on the same chronometer (*e.g.*, a watch she carries in her pocket). But temporal parts must be parts of some Time; so what about the watch that HB is carrying to measure her "own" times?[1]

---

[1] Pp. 108–109. For clarity's sake, I have replaced Mr. Chapman's notation for the two Times of the new framework ('Time$_2$' and 'Time′$_2$') by the simpler 'Time' and 'Time′'.

The first sentence above seems true, but what reason is there to hold the third (beginning with 'Certainly . . . ')? In fact, I am tempted to say that *certainly* it is *not* true. Consider, *e.g.*, the unobjectionable case of bispatiality. Suppose Jones was measuring his back yard when he took the drug that transported him to the other Space (S′). Being a methodical man, Jones continues to measure distances for some time before he realizes that he is not any more in S but rather in S′. Now clearly the distance measured before t (the time when the drug took effect) is *not* spatially continuous with the distance measured after t, although Jones may first think that it is. *Ex hypothese*, the two spatial parts measured by Jones are *not* spatially connected. (There is even no spatial route leading from the one to the other.) Again suppose that after an hour, when the drug's effect wears off, Jones is returned to, roughly, the same spot in his yard. If Jones has been measuring distances all the time he spent in S′, without even realizing that what he was measuring during the last hour was not his yard but some alien space, he would arrive at the surprising (and mistaken) result that his ruler must have shrunk radically between t and t+1.

Betsie's case is no different. In the place s where her velocity reached 5 mach Betsie was transported to another Time, T′. Her chronometer continued, of course, to tick off seconds and moments, but *what* it was in fact measuring west of s was no more T but T′. Suppose again, that Betsie re-entered T at about the same moment (and one mile further west of s) she left it. *If* she were ignorant of the fact that for the last mile she was in T′, she would arrive at the surprising (and mistaken) result that her watch must have run incredibly fast between s and s+1. There simply is no such thing as Betsie's watch's "own" time. A watch is a time-measuring device, and when you take it from one Time to another it measures the other Time. Chapman acknowledges this possibility, but finds it very strange: 'It might begin to measure T′ as soon as HB entered it as if watches can magically know what Time they are in, what time scale is used in that Time and *what* time it is' (p. 109). But surely a measuring device can be used to measure something without that device *knowing* what it measures or what is the scale generally used with respect to the object measured! Jones can use his ruler to measure S as well S′, and (obviously) the *ruler* does not have to *know* which Space it does measure, nor what place it is that it measures. Similarly Betsie's watch most certainly can measure T or T′ without coming to *know* whether *what* it measures is T′ or T.

One more comment seems to be called for. Mr. Chapman makes the interesting suggestion that 'the natural way to interpret the above diagram would not be as a picture of a thing moving from one Time into another but of a world with a *single* two dimensional time.' (The diagram referred to by Mr. Chapman is a chart of Betsie's world-line in Space, Time, and Time′.) However, the idea of a two-dimensional Time is *much* more complex than that of a bitemporal world. A bitemporal world differs from ours only in that the disjunction of temporal relations (earlier than *or* later than *or* simultaneous with) which holds between any two events in our world, fails to hold between some events in the bitemporal world. In a world of bidimensional time, however, any two events (except those forming a line parallel to one of the temporal axes) must be related *both* by one of the above disjuncts *and* by some other, to us unknown, *temporal* ordering relation! I confess that I find this idea extremely hard to comprehend. But even if the theory of bidimensional Time is not impossible, it is clear that it would not apply in the case at hand. The case was of Betsie having uninferred knowledge ("memory") of the events which occurred east of s at all times, yet unable to find any connections between those events and the strange events experienced by her at s. The explanation that the events experienced in s took place in another Time accounts for the lack of connection between the two sets of events. The hypothesis of a bidimensional Time would here be of no use at all, since according to it the events experienced at s *are* connected to the events experienced east of s. In this case there *would* be a causal connection between the strange events and the events experienced east of s partially accounting for the nature of the strange events (when a complete explanation would come only after a survey of the whole temporal plane T-T′ east of s). In our case, however, the conditions were such that a connection between these two sets of events could *not* be discovered. Moreover, to adopt the bidimensionality hypothesis one has to postulate an infinite number of (unexperienceable) events existing at s (and every other place) along the "other" temporal axis, while the bitemporal hypothesis asserts the existence of only two—at t and at t′.

State University of New York at Stony Brook

1 MAR 1974

NOTES

The Analysis Committee consists of: Chairman, P. T. Geach; Secretary, Miss M. A. Boden; J. H. Benson, Alec Fisher, A. Harrison, R. F. Holland, S. Körner, A. C. MacIntyre, A. M. MacIver, B. Mayo, R. G. Swinburne, A. R. White. This committee is responsible for appointing and advising the editor and for the general policy of the paper.

Subscriptions. The subscription to Analysis, Vol. 31, is 24s. (£1.20) in Great Britain and 28s. (£1.40) abroad (U.S.A. $4.00), post free. Each volume comprises six numbers, five of 32 pages and one of 48 pages, appearing within the academic year—in October, December, January, March, April and June. Orders should be sent to Basil Blackwell, 49 Broad Street, Oxford, or placed with any bookseller.

Contributions. Articles submitted for publication should be addressed to Professor Peter Winch, Department of Philosophy, King's College, Strand, London, W.C.2. Contributors are asked to note the following requirements:

Articles should **normally** not exceed 3,000 words in length. **Occasionally,** however, longer contributions can be accepted;

They must be typewritten in double spacing on one side of the paper only;

Footnotes should be kept to a minimum and wherever possible avoided altogether;

Single quotation marks should normally be used, except for purposes of internal quotation and "scare" quotes.

Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

Contributors in the United Kingdom should enclose a stamped addressed envelope of suitable size; if immediate acknowledgment is required a stamped postcard should also be enclosed.

Overseas contributors who wish to have their MSS. returned should send an envelope **and international reply coupons** of the requisite value, whether for air or surface mail.

Proofs are normally read by the Editor. A contributor who wishes to read his own proofs should notify the Editor at the same time as the MS. is submitted.

Offprints. Contributors will be supplied free of charge with a limited number of offprints of their articles. Further copies may be ordered at a charge from the publishers. Orders for these, if required, should be placed **direct with the publishers** when an article has been accepted for publication.

The copyright of articles printed in Analysis remains the property of the author, but contributors are strongly advised, in their own interest, to consult the Editor before consenting to the reprinting of their articles.

PRINTED IN GREAT BRITAIN BY A. T. BROOME AND SON, 18 ST. CLEMENT'S, OXFORD

Vol. 31 No. 3 (New Series No. 140) January 197

ANALYSIS

CALCUTTA UNIVERSITY

Edited by
PETER WINCH

31/1/72

22/9/71

CONTENTS

Characterizing judgments and their causal counterparts — COLIN RADFORD
Virtues and passions — GABRIELE TAYLOR *and* SYBIL WOLFRAM
On killing and letting die — DANIEL DINELLO
Evidence, utility and God — LAWRENCE RESNICK
The principle of mood constancy — ALFRED F. MACKAY
"Jonesese" and substitutivity — A. C. GENOVA
Belief statements and their logic — F. H. GEORGE
Some doubts about illocutionary negation — RICHARD T. GARNER

BASIL BLACKWELL · BROAD STREET · OXFORD

Six shillings net
(Thirty pence)

ANALYSIS 31.3 JANUARY 1971

CHARACTERIZING-JUDGMENTS AND THEIR CAUSAL COUNTERPARTS*

By COLIN RADFORD

1. WHAT is the criterion of what a man wants *qua* desires? Is it, as Russell had said (*Analysis of Mind, III*), what satisfies him, or may it sometimes be, as Wittgenstein had replied (*Blue Book*, pp. 22, 43), what he (sincerely) says he wants? Or, mysteriously, are both criteria employed together in most cases? If a man says, 'Oh—that reminds me', does he not commit himself to a causal, counter-factual claim, *viz.* that, *caeteris paribus*, had not *that* happened, he would not have thought of whatever it was that came to mind? Yet is his remark not an immediate one? And does he not speak here with authority? And does not a similar problem arise *vis-à-vis* a man's statements about his motives for having done something?

These questions are related, as Wittgenstein realized, to another, his discussions of which are reported in *Lectures and Conversations*. Suppose a man to be looking at the façade of a building, or the interior of a room, or simply at a door. There is something wrong, *displeasing*, about its proportions. But what? . . . He may see what is wrong immediately, 'The door is too high', or it may take time for him to come to this conclusion.[1] But, in either case, we can ask what the connection is between the aesthetic judgment and the prediction expressed in the causal hypothetical, 'If the door is lowered, my feelings of dissatisfaction will (as a result) disappear'.

What we have of Wittgenstein's answer to this question is unclear, partly because he appears to hover between at least two answers to it. What I shall call the stronger thesis is that the aesthetic remark, and hence its correctness, is quite—totally—independent of the causal hypothetical (cf., II, §§ 14, 16, 38–9). That is to say, Wittgenstein is here committed to the view that, even if the door were lowered and the critic's feelings of dissatisfaction did not, as a result, evaporate, the aesthetic judgment might nonetheless be correct. Moreover, since there is no relationship of any kind, the aesthetic judgment in no way implies the causal claim.

Wittgenstein was, clearly, very attracted by this view of aesthetics and aesthetic judgments, but it leaves one puzzled and sceptical and with the query, 'What then is the criterion of the aesthetic judgment's being correct? Is there one?'

---

* I borrow these phrases, and the idea for my snow example, from Frank Cioffi who first interested me in the problems discussed in this paper.

[1] We need to imagine that the man is puzzled in order to treat his remark as an explanation, which is what Wittgenstein sometimes calls it (*ibid.*, II, §18).

His weaker thesis is that the aesthetic judgment is connected with the causal judgment, so it has causal implications, but if these turn out to be false that is not sufficient to falsify the original claim (cf. II, §§ 15, 21).

This does not arouse the same feelings of scepticism but, again, it leaves us with the question, 'When, under what conditions, is the aesthetic judgment correct—if it makes sense to speak of correctness here?'

In the rest of this paper I shall try to answer this question, to give a correct account of the relationship between various aesthetic judgments and their causal counterparts, and to explain the basis of these relationships. In so doing I hope to show that Wittgenstein's favoured thesis, though incorrect for some cases is, surprisingly, correct for a whole host of aesthetic judgments.

I shall proceed by examining various examples, not all of which will be aesthetic.

2. To support the last claim I shall consider the following case: I look out of the window at the newly-fallen Illinois snow, snow colder than any snow I have ever seen, and say to my wife, 'Look at the snow. It looks—it looks as if it has been sprinkled with sugar.' My wife agrees.

It may be tedious but necessary to point out that this remark, in this context, does not in any way suggest that the snow has in fact been sprinkled with sugar. If anything, it suggests that it has not—though it may be the context which does this work. But if, surprisingly, the snow had been sprinkled with sugar, that would not falsify the remark or in any way threaten it. Its having been sprinkled with sugar would rather provide a Pickwickian confirmation of our remark: it would similarly confirm, 'It looks as if it *had* been sprinkled with sugar', though this unambiguously implies that it has not.

But now suppose that a third person is present who is of a sceptical, scientific, tough-minded bent. He says, in response to my remark about the snow, 'I don't know about that. Let's see if you're right!' And, taking some sugar, he spreads it over the snow in the garden. As a consequence the snow assumes a dull, glassy look as it slowly begins to melt.

His test shows, perhaps, that I did not know what snow actually looks like when it is sprinkled with sugar. But does it have any tendency at all to show that my remark was incorrect? I think not. His testing of my remark is not just coarse-grained, impolite and socially inappropriate, it betrays a fully-fledged misunderstanding of its nature. (Not that there is any discontinuity here.) For when I said, 'The snow looks as if it has been sprinkled with sugar', I was not—despite grammatical appearances to the contrary—making *any* claim about what snow actually looks like when it is sprinkled with sugar. My claim was what I shall call a characterizing one, *i.e.*, it was a claim which characterized the nature of my experience and, unless I am eccentric, the look of the snow.

Of course, had I known that sugar has a similar effect on snow to salt I might have hesitated to say what I said; it does seem open to an obvious, though vulgar, objection. Yet there are circumstances in which the aptness, happiness of a characterizing remark may be such that we make it despite our knowing that its causal counterpart is false. Thus, we might say of snow that glitters and is very powdery, 'It looks as if it has been sprinkled with salt'. (In England table salt crystals are smaller than those of table sugar.)

So the falsity of this characterizing judgment's causal counterpart does not falsify or in any way threaten its being correct. Moreover, no claim is implied, even in a weak fashion, about how snow actually looks when it is sprinkled with sugar.

But would the *truth* of this characterizing judgment's causal counterpart establish or help to establish its truth? What does this mean? Well, if snow of a familiar temperature and texture, when sprinkled with sugar, assumed the appearance of the strange, unfamiliar snow in my Illinois back-garden, wouldn't *that* corroborate my judgment?

If it does do so, it does so fortuitously. That is, it is only if snow sprinkled with sugar looks as if it has been sprinkled with sugar, *i.e.*, looks as it would look to someone ignorant of the surprising effect that sugar can have on the appearance of snow, that the correctness of the causal counterpart corroborates the characterizing judgment. And, of course, it does so parasitically. That is why the 'confirmation' is Pickwickian.

3. If the preceding claims are correct, what then is the criterion for the correctness of such aesthetic judgments? Is there one?

I shall say immediately that, at least for the above and similar cases, *i.e.* where the aesthetic judgment is a characterizing judgment, the criterion is assent. What, for example, is the criterion for the correctness of the claim that music in a minor key is grave, sad, plaintive, *etc.*? My judgment is correct if others accept it, and accept it because it seems to be apt, to capture and convey the look of the snow.[1]

But the tough-minded man did not accept my remark. Well, he *may* not have done, and we understand that there are insensitive people about. In any case, he may have accepted it alright but is not terribly interested in such remarks. He is more interested in their causal counterparts. Some people have the knack of making such aesthetic judgments. They are born critics. Others have the capacity not merely to characterize, to put into words, our common responses, but to express responses which, when we hear them, become our own. (A friend, describing an impending thunderstorm, in a letter wrote, 'The sky looks bruised and the cars in the car park have a dull glint'.) These are the poets. And we who relish such remarks are their publics. Of course a residue remains; but it is a *fact* that many of us do make such remarks and, on many occasions, they find a general, immediate, acceptance. It is this fact, this concensus of judgment, that makes possible the objective idiom of such remarks and which encourages me to characterize my experience, how the snow looks to me, in terms of how the snow looks. And, all over again, how it looks logically depends on what people say and not on what happens when sugar is sprinkled on snow.

---

[1] In this respect, such judgments are like certain critical judgments which *embody* causal claims; for example, 'Blake's "The Tyger" is a much weaker poem if his spelling is modernized to "tiger" '. This claim is correct if, and only if, people agree, and they do. And although they agree without perhaps actually conducting the experiment of re-writing the poem, if the poem is re-written it does strike them as weaker, but, if it did not, they would have to admit that they were mistaken.

To anticipate an objection, I am not saying that this characterization would be at all likely to occur to us, or strike us as happy, if snow always melted any substance on which it was sprinkled, or if sugar itself were not white, crystalline and most frequently encountered by most of us in fine granular form. Characterizing judgments all have a causal basis, no doubt, however remote;[1] but they do not—always—have causal implications.

But is it not possible for a characterizing judgment to be right idiosyncratically, *i.e.*, to characterize correctly the experience of an individual, even though that experience is eccentric?

It might be tempting to say that, though we do indeed make such judgments, we can give no sense to the idea that they could be correct or incorrect. For in such cases what a man sincerely says is just what he feels. But this view is too simple. For we can fail to characterize our odd feelings about something, fail to capture them in an apt phrase, and we describe such attempts as failures, near misses, *etc.* Moreover, we must all be familiar with this sort of experience: asked if our hair-cuts or the suit being fitted is right—a personal matter—we say that it is, but someone suggests a further alteration, it is made, and then we say, '*Now* it's right!', and we recognize that our prior judgment was *strained.* We had, as we say, 'persuaded ourselves' that it looked right to us when really it did not.

A different case arises when, in retrospect, we judge our impression to have been—not at all strained, but—untutored, imperceptive, immature. Our past judgments correctly characterized our past responses, *e.g.*, to sweet wine, but now we view them as defective in certain ways.

Not all changes of taste, either public or personal, are seen as corrections of earlier judgments. Past fashions sometimes strike us as ludicrous, and we could not now wear them, but we do not necessarily see our earlier judgments as, therefore, wrong—though we may do so. (Surely we were wrong about platform-soled, high-heeled, sling-backed, peep-toed sandals? But we were not wrong to like the New Look.)

---

[1] What is the causal basis of such remarks as, 'It's as if Wednesday were dark red'?

It might be thought that, on the view presented earlier, the correct characterization of an idiosyncratic response must itself be seen as incorrect since, by hypothesis, it would not elicit general assent. But an idiosyncratic response can sometimes be seen as simply that, questions of correctness do not always arise in connection with responses; and the statistically odd responses of a man can be seen as superior to those of his fellows, and for reasons which can be laid out. However, in the sort of cases in which I am interested, *viz.* comparatively simple and immediate responses where there is little, if any, room for expertise, scholarship, and, hence, confusion, when the question of correctness, aptness in a public sense arises—'Does it *really* look as if it has been sprinkled with sugar?'—it can only be settled by reference to how most people respond. (Think here of colour-judgments.)

But it may still be objected that what someone who says, 'The snow looks as if it has been sprinkled with sugar' means is that if sugar were sprinkled on the snow and it did not produce a melting effect, the snow would look as it does look, *viz.* like *this*. That is, the judgment is correct if, and only if, the snow has the same sort of look which other things have when they are sprinkled with sugar—and no melting occurs. This really is an objective matter, *i.e.* independent of what people happen to say. Perhaps the suggested criterion is right when the characterizing judgment cuts across senses, *cf.* 'Reading his book is like wading through slowly setting cement', but it is wrong at least for the example in question.

Reply: The list of provisos, *viz.* that the things do not melt, is clearly too short. Some things might turn the sugar blue, or take on a smoky hue, *etc.* What, then, the objection is trying to say is that the judgment is true if, and only if, the snow has the same sort of look as certain other things have when they are sprinkled with sugar—table tops, for example, or cakes, which are then termed 'frosted'.

Now this is so; but what logically determines whether or not the two things look the same? A patch of blue sky seen through a hole in a white cloud looks a darker blue than the surrounding sky though they have the same hue. Despite this objective fact they do not look the same colour and would only do so if human beings were so constituted that they would sincerely say that they did look the same colour. But they do not say this, and this too is an objective matter but one comprised of concurring judgments.

4. What has been said so far can help us to understand—and enjoy—

remarks which heretofore may have struck us as exciting but problematic. We may have worried unnecessarily about pretentiousness when there was no pretence—even though the author himself may not have been able to give a correct account of the sort of thing that he was saying.

Consider a passage taken from the essay, 'Some lost copies after originals by Giorgione', by Berenson; it provides a splendid example. The author is speaking of a portrait by an unknown hand of an unidentified woman. He writes:

> The Italian lady in her health and magnificence is before us, restlessly energetic, exuberant, full of interest, full of warming sympathy, with a power of carrying everything before her, a source of life and joy to all who surround her—yet cool-headed, penetrating, and ironical, although full of indulgence.
>
> In so far as my poor words can express it, such is the kind of person who looks out at us from Signor Crespi's portrait... (*The Study and Criticism of Italian Art* (1912), p. 85).

Some will reject this passage as mere self-indulgence; but even those who not only enjoy Berenson's pleasure in the picture but—if only in part—share his response to it may feel embarrassed, and not by the poverty of Berenson's words but by their richness. How can Berenson make such large—and such detailed—claims about an unknown woman on the basis of a mere portrait? It is not just that the artist may not have got a good likeness—and since we do not know who he is we are in no position to make informed guesses about this. Indeed, this is a vulgar, a too obvious, difficulty. The main difficulty is with Berenson's apparent assumption that, from a mere likeness, however accurate, reliable inferences may be drawn concerning not just the sitter's appearance and state of health, but her intelligence, personality, judgment and particular details of her relations with others. If phrenology had been discredited by 1912, what possible excuse or explanation can there be for this exercise? Is it mere *jeu d'esprit*?

The explanation is that Berenson is not making inferences, is not describing the unknown sitter, even if his grammar invites this misinterpretation and even though he may not have fully understood himself. Berenson is describing, responding to—not the unknown sitter, but—the painting and, more specifically, to the lady in the painting. And he should be no more than mildly embarrassed if further scholarship were to reveal that the sitter herself was stupid, dull, diffident.

But why this qualification? Why should Berenson feel any embarrassment whatsoever given the preceding distinction and the earlier account of, 'The snow looks as if it has been sprinkled with sugar'?

The reasons are complex. The picture in question is a portrait, and a portrait—in this genre, at any rate—should be a likeness. It is more—and, hence, less—than a mere likeness, but it is that. And there are certain natural connections between a person's appearance, in particular, his physiognomy, and all those other invisible traits which, for convenience, I shall dub his psychology. These connections are most obvious in certain pathological conditions. If we are presented with a picture of a mongoloid, or an individual in the tertiary stages of syphylis, this should be visible and our judgments about the sitter's psychology better than mere guesses. Whether such connections are very common is a difficult question, though not a philosophical one, but there may well be more that we utilize in making psychological judgments than we are consciously aware of. Still, some of them are patent.

Partly because natural and visible traits are incompletely connected with psychology, portraits are not mere likenesses. The good artist tries to convey something of his subject's personality and, usually, more than is connected with, and so can be conveyed by, his appearance. So he will subtly distort that appearance by adding, for example, a lofty brow, an almost totally conventional sign of intelligence. Or he may add to the stature of his subject as a man and a king by adding to his physical stature, again a conventional device, but one used by artists as different as the anonymous Classical Egyptian bas-relief sculptors and Gainsborough in his portrait of Captain Wade. Again, he may try to convey the spiritual beauty of his subject by rendering him physically beautiful, a device used by both Medieval and Early Renaissance artists, though the former believed that the connection between holiness and beauty was indeed natural, though heaven-sent.

Besides relying on these natural and conventional connections to convey psychology, artists also, surprisingly, have a third at their disposal. Consider, for example, the phenomenon of large feet. So far as I know, there is no natural connection between clumsiness and large feet; and if there is, it may well be a negative one. Moreover, so far as I know, large feet have never been a conventional sign of clumsiness in art—though they are, it is true, part of the clown's professional equipment. But my intuitions tell me that, if the subject of a full-length picture were portrayed as having large feet, we would be inclined to see him as clumsy. And an artist, aware of this connection, though perhaps without being aware that he was aware of it, might reduce the relative size of his subject's feet, not to flatter him but to avoid giving the false impression that his subject was clumsy.

Because of the existence of these connections, and because of what a portrait is, Berenson should be a little embarrassed if his enthusiasm for the lady in the picture proved totally inappropriate to the unknown sitter. Such a contingency would suggest either that he, Berenson, lacks insight or knowledge, and so judgment, in this case, or that the artist he admires lacks technique or integrity. But there are, of course, third possibilities which would save him from embarrassment. One is that the subject was unusual in that her appearance and bearing were completely at odds with and belied by her psychology, and the unknown artist, relishing perhaps the unusual situation of not having to flatter, or simply abandoning psychological accuracy as impossible, just painted what he saw.

These connections, then, not only provide a basis for a concensus of response to portraits but also make it possible for that response to correctly describe something of the psychology of the sitter.[1] They thus save much portrait appreciation from a total aestheticism, *i.e.*, from being an activity in which concensus—among the expert, anyway—is the only test for the correctness, aptness, happiness, of what is said. But Berenson is only just saved from such aestheticism, if, indeed, he is saved at all, as we shall see if we look again at the quote. Remarks of that order can only be aimed at the lady in the picture; if they happen to describe the sitter as well, and even if they happen to match a few independent responses to the portrait, that, we feel, would be no more than a happy accident.

But the anonymity of the sitter and the exuberance of Berenson's response to her portrait make this an extreme case. Now think of portraits of well-known persons, and recall the sorts of remark we make about them. When we look at Velasquez's portraits of Philip IV and we say, 'How crass he looks!' are we not speaking of Philip? When we look at Gainsborough's portrait of Sarah Siddons and say, 'She looks so sad', are we not speaking of Mrs. Siddons? It may turn out that Mrs. Siddons was not sad at the time her likeness was "taken",[2] and that would not mean that we were mistaken in what we said. But we would be *surprised* by this intelligence, if only mildly, and we might say, 'Well, she looks sad in the painting'; and to whom does that 'she' refer to here, if not to the sitter?

If Berenson's remarks about his portrait are not about the sitter, many "psychological" remarks about such paintings are about their sitters. It is just that, if they misdescribe the sitter they are not therefore false. The test of a remark like, 'She has a sad-looking face', whether it is made in front of a portrait or a person, is not whether she is sad but whether that is how the face strikes people. And to say this is not to deny that many such remarks are based on an appropriate fact—in this case, that sad people have a characteristic look—or that if people came to realize that the judgment lacked that sort of basis, what had looked sad to them might change its look—to one, possibly, of repose.

---

[1] If other connections existed they might provide a causal basis for a concensus and give no guide whatsoever to the psychology of the sitter. If no connections existed at all, either there would be no portrait appreciation or it would be a totally personal activity.

[2] If she was, Gainsborough, apparently, was not too aware of it. His recorded comment about her face was 'Damn the nose, there's no end to it'.

5. Having described the relationships of certain characterizing judgments to their causal counterparts, I shall now try to show why those relationships are as they are and in so doing solve the problem that originally puzzled us.

Consider, first, an unsophisticated subject who is shown a series of figures. One is a pair of bent lines. The subject is asked if they look bent and he will say, correctly, that they do. And, unless he is very cautious, or the lines only slightly bent, his characterizing remark will carry the implication that they are indeed bent. He is then presented with the Hering or Wundt illusion and, looking at the pair of cross-hatched, straight parallel lines, he will say, 'They look bent'. Because he is unsophisticated his remark implies that they are bent, and, though they are not his remark is nonetheless correct. For these lines do look bent—even to a sophisticated subject whose knowing remark, 'They *look* bent' carries the unspoken rider, '*but* they are not really bent'. And, again, what he says would be correct even if we had fooled him by presenting him with a figure in which the cross-hatched lines were slightly bulged.

So this characterizing judgment, which it is tempting to describe as "subjective", and which is "immediate" and, in a sense, "apodictic", generally carries an implication that is contingent, factual, objective, or causal; though it need not do so. It does not do so in those unusual circumstances in which we suspect, and it is clear that we suspect, that how things look may be no guide to how things are. In other circumstances our characterizing judgment will carry an implication about how things are, and, of course, it will, generally, be to the effect that things are how they look. But sometimes, as we saw above, it will carry the converse implication. However, these cases too are uncommon and involve some measure of illusion—which explains the 'but' in the sophisticated subject's unspoken rider. And, once again, if what is implied turns out to be false, the characterizing judgment, being that, remains true.

But we must now ask, what is the source of these curious implications? In the case under consideration, anyway, the answer must be fairly obvious. For how would one elucidate the notion of 'looking bent'? Things look bent if they look *as bent things look* when viewed by normal observers from a vantage point not in the plane of the bend of the thing viewed. That is to say, we have to elucidate 'looking bent' in terms of 'actually being bent'. And if things that were bent did not generally look bent, and if things which looked bent were not usually bent, 'looking bent' could not have the meaning it has. It is because this is so that a characterizing judgment, 'It looks bent', generally has the causal implication, 'It is bent', and this is generally true. And, as we have seen, because it is a characterizing judgment, it remains true on those odd occasions when what it implies turns out to be false.

It is not quite correct to say that, 'It looks bent' is "subjective" or "apodictic". This remark is couched in an objective idiom, and whether something does indeed look bent is a matter about which a man—probably with defective eyesight—can be mistaken. But, being mistaken about how things look, he can still be right about how things look *to him*, and it is at this level that we get the remark that characterizes the personal experience, or fails to do so. Thus, a man suffering from, say, irregular astigmatism, will sometimes say, 'It looks bent' or 'It looks bent to me', where we would say, 'It looks straight'.[1] His remark correctly characterizes the nature of his experiences, unless he is confused or strained; it is just that they are aberrant. But there is a similar connection between 'It looks bent to me' and 'It looks bent', to the connection between the latter and 'It is bent'. Generally, how things look to one individual will be how they look to most individuals, and, hence, how they look. For this reason we can—often—get from personal impressions, and judgments concerning those, to the way things are in reality.

An account somewhat like the above, but with suitable modifications for individual cases, would fit a great number of characterizing judgments, including, 'The snow looks as if it has been sprinkled with sugar', and, 'She has a sad face'. But Wittgenstein's example, 'The door is too high', requires another story.

6. I take it to be self-evident that there is, for Wittgenstein's example, nothing which answers to 'The snow has in fact been sprinkled with sugar' or 'It is bent' or 'She is sad'. To think that there is is to imagine that 'It's too high' might point to something beyond the concensus of human feelings about the door, and there is no such thing.

But, clearly, we must look elsewhere for some explanation why, when Wittgenstein was considering the possible implications of his judgment, he did not refer beyond his own subsequent responses to a hypothetical alteration to the door to the contemporaneous feelings of others about the door. His judgment is, as he says, a reaction, it is not a hypothesis, but how others feel about the door bears on his judgment. What views Wittgenstein may have held about the sureness of his aesthetic judgments are irrelevant here. A more appropriate consideration is simply this, that Wittgenstein begins with the view that his reaction to the door is, in some way, apodictic, and so cannot be overthrown. Hence, considerations about how others might feel, though relevant for testing the correctness of his remark as a value judgment, are irrelevant to questions about how or whether, *qua* apodictic response —however idiosyncratic it might be—it can have implications that are contingent, *etc.* (In this way, Wittgenstein's discussion has little to do with aesthetic *evaluation* as such.)

---

[1] Though rarely, of course, for other than visual clues usually bring his judgments into line with ours. Such conditions are usually revealed only by tests using, *e.g.*, an Amsler Grid.

Now Wittgenstein's own favoured view is that, apparently because, *qua* response, the remark is incorrigible, *i.e.* because there is no question about its locating the nature of and target for his feelings about the door, it can have no implications that are contingent. But that only follows if the implications are classical. If they are weak, there is no incompatibility between the two theses, only a problem about how any subjective apodictic judgment can have contingent implications of any sort.

In any case, Wittgenstein's assumption that, *qua* response, the remark is incorrigible, is not quite right in certain circumstances. A man puzzled by what was wrong with the door might hazard that it was too high, and then reject this as soon as he had said it. Or in an argument he might sincerely insist that it was too high, but come later to admit that the claim was forced.

Of course, a change of mind might come about as a result of the door's being altered, *i.e.* through experiment. This does not mean that what might be called a Russellian view of the relationship between the characterizing claim and its causal counterpart is right, even in this sort of case. For the experiment, by itself, could only result in the man's saying something like, 'It's wrong *now*', and his view that really it was O.K. before, *i.e.* that his past judgment was wrong, would not necessitate his agreeing that it had mischaracterized his past feelings about the door—although he might *also* say that. The "Russellian" thesis gets within a hair's breadth of being right in such a case.

In contrast, the view that the remark does not weakly imply that if the door were lowered his feelings of dissatisfaction would disappear *is* correct in certain circumstances. Suppose the proportions of the room were such that if the door were lowered it would still not look right, but too small. If Wittgenstein knew this, and others did, if this kind of eventuality were fairly common and recognizable, the remark would not have the implication it usually has—even though the door *does* look too high. Which is to say that it does usually have this weak implication, and for reasons which, I hope, are now only too familiar.

I think we are now in a position to solve those other problems, akin to ours, which I mentioned at the beginning. But to do this would require another paper.

University of Kent, Canterbury

VIRTUES AND PASSIONS

By Gabriele Taylor *and* Sybil Wolfram

IT has often been suggested that there is a connexion between virtues and passions, but the nature of the connexion has remained obscure. Sometimes it is held that certain passions are in some way necessary to virtues, that for example a man cannot act generously unless he has generous feelings or that true kindness requires kindly feelings. Aquinas thought that justice cannot be without passion for 'no man is just who rejoices not in just deeds' and joy is a passion.[1] Aristotle at least flirts with a more general version of this line of thought when he says:[2]

> A man is temperate who, abstaining from bodily pleasures, finds this abstinence pleasant; if he finds it irksome, he is intemperate. Again, it is the man who encounters dangers gladly, or at least without painful sensations, who is brave; the man who has these sensations is a coward.

Rather more commonly, however, it is maintained that virtues are concerned with controlling or overcoming passions. Aquinas applies this doctrine to only some virtues, that class which includes courage and temperance. In other cases, he thought, there is no more than a contingent relation between the presence of a passion (*e.g.* anger) and acting contrary to virtue (*e.g.* striking an unjust blow). But this view, too, has been put forward in a more generalised form. In discussing the doctrine of the mean Aristotle says:

> It is possible, for example, to experience fear, boldness, desire, anger, pity and pleasures and pains generally too much or too little or to the right amount. If we feel them too much or too little we are wrong. (*Ibid.* Bk. 2. Ch 6)

Kant can conceivably be interpreted as holding the much tougher doctrine that we must not only control our feelings but act against them if our action is to have moral worth. Though not perhaps his considered view, he does put forward as the paradigm case of virtue that of a man who, overcome by sorrow and quite indifferent to other people, yet manages to act benevolently towards them.[3] Probably the clearest and certainly the most fully fledged version of the doctrine that virtue consists in overcoming passion is given by von Wright in *The Varieties of Goodness*. His claim there is that:[4]

> The role of a virtue, to put it briefly, is to counteract, eliminate, rule out the obscuring effects which emotion may have on our practical judgment.

He continues:

---

[1] *Treatise on the Virtues*, Q. LIX. Art. 5.

[2] *Nichomachean Ethics*, Bk. 2, Ch. 3.

[3] Kant *Groundwork of the Metaphysic of Morals*, Ch. 1.

[4] *The Varieties of Goodness* 1963, p. 147.

> In the case of every specific virtue there is some specific passion which the man of that virtue has learnt to master. In the case of courage, for example, the passion is fear in the face of danger. In the case of temperance it is lust for pleasure.

And again (p. 153):

> The considerate man has learnt to conquer the influence of selfish impulses on his judgment.

Our object is to try to see what relations between passions and virtues can be salvaged from this rather confusing medley of views and to show that they are on the whole less significant than has been supposed.

Two theses can be eliminated rapidly, the first on grounds of triviality, and the second on grounds of implausibility.

(1) The trivial thesis: this states that if we experience a passion when called upon to exercise a virtue, and if the passion prompts us to act in a way contrary to the exercise of the virtue, then we have to overcome (*i.e.* act against) the passion in order to act virtuously. If I ought to be working then in order to display the virtue of industry I must overcome any desires I may have to stay in bed or go to a party. Conversely, any passion the following of which would result in doing the virtuous act should be followed if I am to act virtuously. This thesis clearly tells us nothing, and in particular requires no specific connexion between any one passion and any one virtue. All it says is that to do a virtuous act one must act against desires not to do it, and in accordance with desires to do it.

(2) The implausible thesis: according to this thesis, the elimination of passions, or of certain particular ones, or alternatively their cultivation, as such constitutes displaying or possessing virtues. No one would seriously maintain that virtues consist of no more than this. It is doubtful whether it is true of any; certainly it is not true of all of them. That I succeed in not feeling fear or in feeling benevolent towards others could not possibly be sufficient for me to be said to have displayed or to possess the virtue or courage or benevolence. *Doing* things is plainly also relevant. If one does not do the dangerous thing when one should or does not help those in need then one hasn't been courageous or benevolent whatever the state of one's feelings. Hence subduing or fostering passions, however interpreted, cannot be sufficient for virtuous behaviour. At most it could be necessary.

We are left with two more convincing theses. One (thesis 3) concerns the application of virtue terms to particular actions, the other (thesis 4) to persons, *i.e.* one relates passions to individual displays of virtues, the other to their possession.

Thesis (3): It is plainly unconvincing (*pace* Aristotle) to suggest that one must necessarily have certain feelings (gladness, pleasure?) in order to display virtues like courage or temperance. Equally incorrect, though of greater popular appeal, is the converse suggestion that such virtues as generosity or honesty require eliminating or acting in spite of specific passions, or indeed any passion at all. As we have tried to show elsewhere, what makes an action an instance of an other-regarding virtue like generosity or honesty is not the reason there may be against doing it, and what makes an action an instance of a so-called self-regarding virtue like courage or temperance is not the nature of the reason for which it is done.[1] If so, then no particular passion could need to be overcome in order to act generously, nor any particular one cultivated if one is to act courageously.

But it seems at least more plausible to hold that:

(i) in order to display a self-regarding virtue on any one occasion a man must necessarily overcome (act against) some passion, and which virtue of this class he displays will depend on which passion he is acting against;

(ii) in order to display an other-regarding virtue on any one occasion a man must necessarily act in accordance with some passion, and which virtue he displays will depend on the nature of the passions he follows.

The apparent plausibility of this view rests on the fact that in order to act *e.g.* generously or conscientiously one must have a particular sort of reason for one's action, viz. to confer a particular kind of benefit on some other person, while it is the nature of the reason against the action which identifies it as an *e.g.* courageous or prudent one. And there is a certain resemblance between the relevant reasons and passion which makes it tempting to equate them, particularly in the self-regarding case. For when a man displays a failing of this class, as when he yields to passion, he acts contrary to Reason in that he does what he thinks it would be better not to do for the sake of a short-term and generally immediate gain. Being imprudent, for example, involves taking an immediate advantage or pleasure at the cost of some more substantial but possibly more distant one, being cowardly, purchasing safety at the cost of a more lasting benefit. And on the other side it may seem that generous feelings are an essential part of generosity. After all, a man who does not give gladly, who perhaps begrudges the time or money spent, may diminish or even destroy the benefit he hopes to be conferring, for we do not normally enjoy what is given merely out of duty.

But attractive though this may seem at first sight, it is easy to see that it is in fact a mistake to equate with passions the reasons in terms of which the virtues themselves are identified. There are two objections to doing so:

1. First and most important, it is only in the exercise of a very few virtues that we have a plausible passion at hand to be either overcome or

---

[1] Taylor and Wolfram 'The Self-regarding and Other-regarding Virtues', *Philosophical Quarterly*, vol. 18, 1968. 'Self-regarding' is a misnomer, but we retain it for convenience.

followed. Even if to act courageously or in a temperate manner involves overcoming fear or lust, there are no such suitable passions in the case of other self-regarding virtues, such as industry or prudence. We do not have a single passion for not working or for having an immediate advantage, but just possibly a passion for staying in bed or going to parties which leads to not working when we think we should, or for this or that which would mean a loss of long-term advantage. Nor again do we have a single passion for exhibiting our passions which has to be overcome if we are to exercise self-control, but only a variety of passions any of which we must refrain from exhibiting when it seems better to conceal it. However one may wish to classify the passions, or whichever traditional list of passions one may consult, it seems plain that their classification does not coincide with the classification of the initial reasons against action which serve to identify virtues like these. The man who wants to spend all his money at once on a Rolls Royce and the man who wants to give it all to charity now both have the sort of *prima facie* reason which, if (wrongly) acted upon, makes their acts instances of the same failing, *viz.* imprudence. But the respective passions which prompted them (supposing they were so prompted) would not ordinarily have the similarity necessary to make them candidates for being the same passion.

The difficulties are even greater in the case of the other-regarding virtues. Those advocating thesis 3 have never attempted to specify the passions which are supposed to be required when exercising *e.g.* justice or conscientiousness; they have confined themselves to the remark that a man must experience gladness or joy or pleasure. But if the passions are left indeterminate then evidently they cannot be used to identify the virtues. Perhaps the thought is that to act generously one must have generous feelings, to act kindly, kindly ones, to act justly, just ones. But this would be to admit that there just is no list of passions corresponding to such virtues, for such a step amounts to identifying the passions by prior identification of the virtues, whereas advocates of thesis 3 want to identify the virtues by means of the passions.

2. Secondly, even in the most favourable cases it seems clear that reasons for acting in a certain way cannot be identified with experiencing certain passions. 'Perceiving the danger' cannot be equated with, or even said to entail, 'feeling fear', nor can 'expecting a pleasure' with 'feeling lust', let alone 'belief that one is conferring a benefit on others' with 'experiencing feelings of generosity or love'. I may *e.g.* perceive the danger and not feel fear because the emergency requires me to act so swiftly that I have no time to feel anything, or because I am in the grip of some other passion which leaves me insensitive towards danger, or again because I am well practised in warding off fear. Equally, I may expect pleasure and yet feel no cravings. Given this, the thesis that, for

instance, exercising courage necessarily requires acting in spite of fear loses whatever plausibility it may have had: it seems absurd to suggest that a man who faces danger in the belief that it is worthwhile to do so is not acting courageously just because as it happens he does not feel fear. Thesis 3 also implies that the degree of *e.g.* courage a man displays depends on the extent of his fear, and this is equally absurd. It is not the amount of fear he is experiencing but the risk he believes himself to be running that is relevant: how probable it is that harm will come to him and how great the harm is likely to be, given his specific circumstances. If A thinks he may well drown in swimming across the river and B that he might catch cold, then short of the cold being unusually disastrous, A will show greater courage in plunging in than B, however weak or strong their respective feelings of fear may be. Similarly, however passionate my cravings, I cannot be said to have displayed a great deal of temperance if the pleasure I forego must be regarded as a minor one, given my beliefs about the situation. Nor can I be said to act very generously if my gift is on the mean side, however filled with generous feelings I may be.

There is, then, nothing to be said for thesis 3 as it stands. But two empirical points, albeit not of very great interest, can be extracted: (i) By and large it is likely that we feel more fear the greater we believe the danger to be, more passionate cravings the greater the pleasure in view. In such cases it is of course correct that one must act in spite of one's great fear and passionate craving in order to be courageous or temperate, and in such cases (*i.e.* given that the great fear is caused by great danger) it is also true that one's acting in spite of such fear is an indication of having displayed a considerable amount of courage. (ii) In order not to spoil his generous or kind act, a man must disguise any reluctance he may feel and at least appear to give *etc.* gladly. But the sorts of acts that can be spoilt by being done plainly "out of duty" are limited. While we may be sensitive about how *e.g.* gifts are given, we are not nearly so concerned when it is a question of being dealt with honestly or fairly. And a judge who displayed gladness in sentencing a criminal to death would simply be adding insult to injury.

Thesis (4). It might be the case that possession of virtues is more closely linked with overcoming or cultivating passions than their exercise seems to be. And it seems possible to extract two initially not too implausible theses along these lines. Although rather different, both can be seen as arising from the fact that for someone to be said to possess a virtue he must display it, and not the corresponding failing, in all (or if we are not too particular, most) relevant situations.

1. It might be said that to possess virtues a man must in some sense have mastered his passions. Although, as we have seen, the state of someone's passions is irrelevant to the classification of his acts as courage-

ous or generous, it is of course true that he cannot be said to possess these virtues unless he displays them even when he experiences passions running counter to their exercise. He must therefore have mastered such passions in the sense that he is able either to eliminate them at will or to act in spite of them; otherwise he may fail to face dangers or to part from his possessions when he feels frightened to do so, miserable at the thought *etc*. Similarly, he must be able either to induce, or to act in the absence of, passions for doing acts which are instances of virtues; otherwise he will fail to act as he thinks he should on occasions when the course of action in question does not itself induce such feelings in him.

This harmless and not very interesting variant of the trivial thesis is apt to be confused with other more interesting but less harmless ones, in particular: (i) that possession of each virtue requires the mastery of some specific passion, and (ii) that possession of a virtue requires a man to have "mastered passions" in the sense of having actually acted in spite of one. But these, unlike the initial statement, are at best rather doubtful contingent points.

(i) To possess the virtue of industry or generosity one must, according to the thesis in its harmless form, have mastered any passion which interferes with its exercise and so, *inter alia*, a passion for lying in bed or love of one's possessions. But while these may be passions which in fact commonly interfere with the exercise of these virtues, they have no logical priority over other passions, nor indeed is it obvious that mastery of other passions is not in practice of equal importance: a passion for parties or for gambling might well stop a man working, and envy or hatred of the recipient prevent him from acting generously. There is in most cases not even empirical justification for demanding mastery of one specific passion rather than anyway some others for the possession of this or that virtue.

(ii) has a logical point in its favour, *viz*. that exercising a virtue when he experiences passion P or does not experience P1 is proof, probably the only proof, of someone's ability to act contrary to P or in the absence of P1. But it claims more than this. For possession of a virtue is supposed to be a state of some duration, and it is clear that having acted contrary to P on one occasion can be no more than *evidence* of ability to do so on subsequent ones. How good it is as evidence depends on the strength of the correlation between being able to act contrary to P on occasion 1 and being able to do so on occasion 2, and this is likely to depend among other things on what P is. In particular, even if it is true that acting contrary to P on some occasion is good evidence of future mastery of whatever is reasonably considered the same passion (*e.g.* fear, love of one's possessions), it does not follow that it is equally good evidence of mastery of any other passion, let alone of all others or all those likely to interfere with the exercise of any particular virtue. That a man has desisted from gambling when he felt a passion to do so may be good evidence that a passion for gambling will not interfere with the exercise of virtues; but it is clearly not good evidence that no passion will ever make him fail in *e.g.* industry or thrift. Moreover, even where there is a strong correlation between being able to act contrary to P on one occasion and being able to do so on another, so that a man can be said to have 'mastered P' on the strength of one such occasion, it does not follow that this is the only or best guarantee of such mastery. Techniques such as counting to ten or fixing one's mind on something else may afford reliable methods of preventing the onset of fear or anger; if so, a man's acquisition of such techniques affords as good a guarantee that he will not fail in courage or patience through such passions as the fact that he has in the past done the courageous or patient thing in spite of feeling them. Hence practice in facing situations which induce passions is not, as Aristotle seems to have thought,[1] necessarily the only or best way of achieving mastery of passions in the sense in which such mastery is a necessary condition of possessing virtues.

2. There is again a minor truth, though this time of a wholly contingent nature, in the thesis that the agent should find the exercise of a virtue a pleasure or at least not irksome to be said to possess it. We have seen that the agent's enjoyment or otherwise is no measure of the virtue of his acts, nor is his pleasure in them even something we always care for. In one sense the same is true of the possession of virtues. Someone who regularly practises a virtue must be said to possess it, and to a high degree if he does so in difficult as well as easy situations, no matter what he may feel. Again, only in some cases do we account it a merit in a man to take pleasure in the practice of a virtue, namely, where his pleasure derives from benefits he is conferring on others; we do not like him to derive pleasure from any harm to others which may be involved in the exercise of *e.g.* justice or courage. And the reason for this seems to be not that the former is in itself a virtue or essential to virtue, but rather that someone who strays from the path of virtues to indulge his pleasures is less likely to cause harm to the world at large if it is conferring benefits and not injuries which he enjoys. There is however this much to be said for the present thesis, *viz.* that possession, or secure possession, of a virtue requires that it should be practised without difficulty, and how "gladly", *i.e.* easily, a man practises a virtue is some sort of measure of how well-established that virtue is. Hesitation in doing the just or courageous thing when it is worth doing is a sign of poorly established justice or courage. Similarly, that someone craves for cigarettes suggests that his temperance in this matter is not very well confirmed. This is probably not a logical but a contingent truth. The point is not that 'A's courage is well established' means 'A always performs courageous acts

---

[1] *Nich. Eth.*, Bk. 2, Ch. 2.

without a struggle', but rather that empirical evidence suggests that if he has a struggle to do such acts he is likely sometimes to fail, and if he finds difficulty in facing relatively easy situations, he is the less likely to be able to manage more difficult ones.

Thus the only correct logical point which thesis 4 seems to offer us, or indeed we are left with in the end, is that overcoming any passions which interfere with its exercise is a necessary condition of being said to possess, as it is of being said to display, a virtue. And it is important to notice that this is no more than a necessary condition. A man whose passions never led him astray might nevertheless lack at least all the other-regarding virtues since, although never governed by passion, he might be motivated solely by self-interest.

University of Oxford

ON KILLING AND LETTING DIE

By Daniel Dinello

JONATHAN BENNETT in his paper 'Whatever the Consequences' (Analysis, 26. 3) attempts to refute what he refers to as the conservative position on the following problem:

> A woman in labour will certainly die unless an operation is performed in which the head of her unborn child is crushed or dissected; while if it is not performed the child can be delivered, alive, by post-mortem Caesarian section. This presents a straight choice between the woman's life and the child's (p. 83).

The conservative position is as follows: The child's death is part of a killing; but, in the case of the mother, there is no killing and death occurs only as a consequence of what is done. Therefore, the principle, 'It would always be wrong to kill an innocent human being whatever the consequences', when added to the premise, 'operating involves the killing of an innocent human being', yields the conclusion: 'it would be wrong to operate'.

Part I of this paper is a brief exposition of Bennett's attempt to refute the conservative position; Part II consists of two counter-examples to Bennett's position; and, Part III is my analysis of the issue.

I

Bennett states correctly that, without an appeal to authority as ground for the principle, the conservative must argue that the premise: 'In this case, operating would be killing an innocent human while not-operating would involve the death of an innocent human only as a consequence' gives some reason for the conclusion. The conservatives have drawn the action/consequence distinction correctly, namely, operating is killing while not-operating is not. The questions are: By what criteria is the distinction drawn and are the criteria morally significant in this case?

Bennett argues correctly that a number of criteria could support a moral conclusion, but are irrelevant in this case. One criterion remains, namely, not-operating is not killing-the-woman because it is not doing anything at all, but is merely refraining from doing something. This is the conservative's final support. The question now is: Is there any moral significance in the acting/refraining (*i.e.*, killing/letting die) distinction?

Bennett suggests that the conditions for distinguishing between 'x killed y' and 'x let y die' are the following:

(1) x kills y if (a) x moved his body
    (b) y died
    (c) there are relatively few other ways x could have moved which satisfy the condition: if x moved like that, y would have died.
(2) x lets y die if (a) x moved his body
    (b) y died
    (c) almost all the ways x could have moved satisfy the condition: if x moved like that, y would have died.

Bennett concludes that since the conservative position rests on there being a morally significant difference between killing and letting die, and since there is no moral significance in the distinction based on the number of moves the agent can make, the conservative position has absolutely no moral bite.

II

The following two counter-examples show that Bennett's conditions for drawing the 'killing/letting die' distinction are incorrect.

Case I: Jones and Smith are watching television. Jones intentionally swallows a quantity of poison sufficient to kill himself. Smith, who knows the antidote, pulls out a pistol, shoots, and *kills* Smith. But, according to Bennett's criteria, this would be a case of 'letting die' since almost all the moves Jones could make (*i.e.*, moves other than, *e.g.*, forcing the antidote down Smith's throat) would satisfy the condition 'if Jones moved like that, y would have died'.

Case II: Jones and Smith are spies who have been captured by the enemy. They have been wired to each other such that a movement by one would electrocute the other. Jones moves and kills Smith. But, according to Bennett's criteria, this too would be a case of 'letting die' since almost all the moves Jones could make, *etc.*

Bennett's conditions for drawing the distinction are clearly wrong, but it remains to be seen whether his conclusion ('the conservative position has no moral bite') is correct.

III

The following are what I take to be the conditions which distinguish 'x killed y' from 'x let y die':

(A) x killed y if x caused y's death by performing movements which affect y's body such that y dies as a result of these movements.

(B) x let y die if (a) there are conditions affecting y, such that if they are not altered, y will die.
  (b) x has reason to believe that the performance of certain movements will alter conditions affecting y, such that y will not die.
  (c) x is in a position to perform such movements.
  (d) x fails to perform these movements.

The following are clarification and justification of these conditions:

(1) Part (b) is necessary, in that we would not want to say that a person who knew no way of altering conditions that are affecting y had let y die. For example, suppose y is dying of an incurable disease and a doctor, x, has no choice, but to watch y die. It would not be true that x let y die.

(2) (c) is necessary because the other conditions could be fulfilled, and if y were incapable of performing the movements, we would not say that he had let y die. For example: y is dying. X knows what movements would alter the conditions affecting y, but he has been securely tied to a chair.

The 'killing/letting die' distinction drawn in terms of the number of moves the agent could make clearly can have no moral significance. It is not obvious, though, that the distinction as I have now drawn it could have no moral significance. Consider the following example: Jones and Smith are in a hospital. Jones cannot live longer than two hours unless he gets a heart transplant. Smith, who had had one kidney removed, is dying of an infection in the other kidney. If he does not get a kidney transplant, he will die in about four hours. When Jones dies, his one good kidney can be transplanted to Smith, or Smith could be killed and his heart transplanted to Jones. Circumstances are such that there are no other hearts or kidneys available within the time necessary to save either one. Further, the consequences of either alternative are approximately equivalent, that is, heart transplants have been perfected, both have a wife and no children, *etc.* On Bennett's analysis, there is no morally significant difference between letting Jones die and killing Smith: the consequences of either alternative are equivalent and there is no moral distinction between killing and letting die. But, it seems clear that it would, in fact, be wrong to kill Smith and save Jones, rather than letting Jones die and saving Smith.

Further, suppose that Jones and Smith are in the same situation, but there is one difference between them: Jones has a wife and Smith does not. Bennett would have to say that since killing and letting die are morally distinguishable only by reference to the consequences of each alternative, the doctor ought to kill Smith and save Jones (Jones' death would sadden his wife, but Smith has no wife and other things are equal). But, this also seems to be wrong.

Bennett argued that the conservative has absolutely no morally relevant factor to which he could appeal, *i.e.*, the 'killing/letting die' distinction is not morally significant. The preceding two examples show this conclusion to be false: There are cases where consequences are equivalent and cases where the consequences of killing are preferable, yet still wrong to kill. The distinction as I have drawn it has some moral bite: it seems intuitively clear that causing a death is morally somewhat more reprehensible than knowingly refraining from altering conditions which are causing a death. Bennett has not refuted the conservative position because the question of whether an act is one of killing or letting die *is* relevant in determining the morality of the act. The conservative, though, gives this factor absolute status. In order to refute this position it must be shown that in many cases other factors outweigh this one. In other words, the question is not whether the conservative position has moral bite, but rather how much moral bite it has.*

---

* I am indebted to David Blumenfeld (University of Illinois, Chicago Circle) for criticisms of earlier versions of this paper.

University of Wisconsin

EVIDENCE, UTILITY AND GOD

By LAWRENCE RESNICK

IN the March 1970 issue of this journal K. G. Gotkürdel and J. King-Farlow (G and K-F) argued that (a) the following propositions are not formally contradictory:

(1) Evil Exists.
(2) A perfect (that is, a completely good and omnipotent and omniscient) Creator exists.
(3) A perfect Creator would only allow Evil to exist if He had a morally sufficient reason (such that all Evil served a good purpose in the best way).
(4) Appallingly much evil is '*prima facie* gratuitous', that is, appears to us to serve no good end whatsoever.
(5) There must, therefore, be a morally sufficient reason for Evil's existence, but in many, many cases the reason remains utterly mysterious to us.

(b) A fairminded thinker, having satisfied himself that there is no formal reason for *not* believing the set of propositions (1) to (5), is free to assess independently the probability of (2) and the utility of believing (2); and (c) if he finds evidence for (2) in the ' "teleologically" pleasant features of the world' and utility in believing (1) to (5) then, since these considerations are rationally respectable, he cannot fairly be accused, on the grounds of believing both in God and Evil, of being rationally deficient.

G and K-F presuppose that 'Evil exists' and 'An all Good, all Knowing, all Powerful God exists' are not logically incompatible. For the purpose of this discussion I shall accept that presupposition also. Consider this parallel series of propositions:

(6) Every person I meet is single-mindedly engaged in a plot to destroy me.
(7) Judging by outward appearances, most people I meet are indifferent to my presence and some are concerned for my welfare.
(8) People who are single-mindedly engaged in a plot to destroy me would be apparently indifferent or helpful only if their behaviour contributed in some way to my destruction, for example, by lulling me into a false sense of security.
(9) Therefore, the apparently indifferent or helpful behaviour is part of a plot to destroy me—a plot the details of which are unknown to me.

Clearly, (d) propositions (6) to (9) are not inconsistent; (e) even a paranoid might have real enemies whose existence would provide some independent evidence for (6); and (f) to a paranoid, belief in these propositions has the highest utility for survival (and his psychiatrist might say that the paranoia has utility as a defence against personal guilt). However, since none of these considerations tends in the least to show that this set of beliefs should be accepted as rational, I conclude that there are grounds for suspecting a flaw in the form of argument used by G and K-F. A detailed examination follows.

Rationality. It is unhelpful, I think, to characterize the problem of Evil as a problem about the rationality of theists. Suppose that the existence of evil *is* incompatible with the existence of God, and suppose that a serious man with powerful inclinations toward theism said, 'I *see* the contradiction—or at any rate the logical side of me sees it—but I cannot bear to contemplate the world and all of its gratuitous evil without believing that somehow it is necessary and good. I must believe in God. Some demands are more powerful than a commitment to consistency.' Now it seems to me quite inappropriate to say either that *he* is irrational or even that his belief is irrational. It is certainly, by hypothesis, *unscientific*, *unwarranted* on the basis of logical reasoning, and in that narrow sense *non-rational*. But what we normally mean by rationality has a deeper basis in action, I suspect, than is borne by a belief in God, *per se*. In any case I prefer to put aside the question of characterising theists by stating the problem of Evil in terms like this: Is 'A perfect Creator exists' logically incompatible with 'There is evil in the world', and if not, does the evil in the world rule out postulating the Creator God as a reasonable explanatory hypothesis? Since, to repeat, it is a presupposition of this discussion that 'God exists' is logically compatible with 'Evil exists', I move directly to a consideration of the second question.

Independent Evidence. G and K-F's claim that the 'teleologically pleasant' features of the world provide independent evidence for the hypothesis of a perfect Creator is mistaken in much the same way that my claim (e) is mistaken. (So I withdraw (e).) They are saying that although the world contains appallingly much gratuitous evil which counts against the hypothesis of a perfect Creator, it also contains, just as undeniably, some good things which must be admitted as evidence *for* the hypothesis. But this would be like saying of a tall English pacifist who visited the Isle of Elba and had a wife named Josephine that there was independent evidence that he was Napoleon. Once an hypothesis has been ruled out by the available evidence no part of that evidence can be construed as independent evidence for the hypothesis unless the inferences used come under challenge, and that is not in question here.

If *all* one could see were the teleologically pleasant features of the world and if the *only* hypothesis one had was that there was a perfect

Creator (or if all other hypotheses were inconsistent with the pleasant features) then perhaps one could claim to have independent evidence of a perfect Creator. Even this much is by no means clear, for the initial plausibility of the Creator hypothesis depends to a great extent on the evidence of design, and we must therefore infer the character of the Creator from the character of the design. But what could ever lead us, acting independently of the religious tradition, to postulate, even while attending only to the best features of the world, an unlimited, *perfect* Creator? I can imagine being overwhelmed by some beautiful feature of the world—a mother's love for her child—and saying, 'The Creator must be perfect!', but I cannot imagine that continued observation of such phenomena would support that statement as a serious hypothesis. The characteristic of perfection would have to be eliminated. In any case one cannot help seeing many ugly features of the world as well and since those data are clearly incompatible with the hypothesis of a perfect Creator, it makes no sense to speak of the pleasant features of the world as independent evidence of a perfect Creator.

Utility. G and K-F give the following example to show the part played in their argument by the concept of utility.

> A cook has made a barely adequate omelette from eleven of the twelve eggs available, but has some reasonable doubt as to the freshness of the last egg (p. 143).

Even if the last egg is fresh the most it could contribute is a nine per cent increase in the bulk of the omelette (assuming the eggs are all the same size) whereas if it is spoiled it might make the omelette one hundred per cent inedible. Therefore a reasonable man, given only the slightest evidence that the last egg is spoiled, would leave it out of the omelette (assuming he believes that all the other eggs are fresh). G and K-F argue that analogously, given the (admittedly slight) independent evidence of a perfect Creator, a fair-minded man who holds that 'the utility of believing . . . proposition (2) and being right about the "eternal things" is staggeringly great', can plausibly conclude that because the utility is so high, all he needs to justify believing in a perfect Creator is some slight evidence, which he has.

But first, as I have just shown, there is no independent evidence at all of a perfect Creator. Second, the utility to which G and K-F are trying to appeal arises from the consequences of *acting* on the belief, rather than from the psychological benefits of holding to the belief itself. The point of the egg example is that the cook finds it inconvenient, because he is pressed for time, or difficult, because he has a cold and cannot trust his sense of smell, or impossible, because it seems to be a borderline case, to determine with assurance whether or not the egg is fresh. Nevertheless he must act immediately as if the egg is clearly fresh, or alternatively, as if it is clearly not. Considering the risks involved each way, he wisely decides not to use the egg. This is simply not parallel to the question of believing in God. First of all, there is no reason to doubt the status of the latter, taken as an explanatory hypothesis. It would be nothing short of a perversion of the scientific method to conclude from the available evidence that there is an all Good, all knowing, all Powerful Creator. Second, how can the utility of acting on that hypothesis be assessed when, in fact, it is far from clear what would count as acting on it? But even if all theists acted in some particular way which had public consequences of great utility it would not follow that theism was justified as an explanatory hypothesis.

Consider the following case. The cook, a nervous man given to timidity, buys eggs from his brother-in-law, a large brutal man who is the fraternal twin of the cook's wife. Recently at least one egg in a dozen has been rotten and, fortified by a stiff drink of cooking sherry, the cook found the courage to threaten to buy from another source if he found one more rotten egg. The twins were furious but visibly more respectful. Now, as the cook breaks open the last egg, evidence of his senses suggests that it is very likely inedible. However when he takes into account the fluttering in his stomach, the emptiness of the sherry bottle and his intense premonition of disaster, he decides that since the utility of avoiding a confrontation is staggeringly great, he will believe that the egg is fresh—and he uses it in the omelette.

My sympathy for this harassed man cannot extend to the point of saying that his *belief* that the egg was fresh was correct even if I admit that, given the circumstances, his *action* was justified. Suppose that, for some external reason, no one ate the omelette and that the result of the cook's action was that he and his wife lived happily ever after. The evidence against the freshness of the egg cannot be overwhelmed in any degree by the benefits to the cook, however great, of ignoring the evidence. Similarly, even if a widespread belief in a perfect Creator would increase domestic tranquility and promote the general welfare—outcomes devoutly to be wished—this would not overwhelm, in any degree, the more than sufficient grounds for rejecting the hypothesis of a perfect Creator as an explanation of the physical universe.

There are more realms in human nature than are revealed in the construction and evaluation of deductive and inductive arguments and I should prefer to see attempts to explain the legitimacy of theism which take account of it as a religious doctrine and avoid the vocabulary of *utility*, *evidence*, and *hypothesis*.

State University College of New York at Old Westbury

THE PRINCIPLE OF MOOD CONSTANCY

By ALFRED F. MACKAY

IN the final paragraph of my 'Inferential Validity and Imperative Inference Rules' (ANALYSIS, 29.5) I suggested that perhaps the most promising alternative for showing the non-divergence of imperative from indicative logic might be to retain Hare's "speaker-commitment" conception of validity for imperative inferences (including mixed inference) while developing some other special principles governing such inferences than Hare's two rules. Professor David Clarke, in his 'Mood Constancy in Mixed Inferences' (ANALYSIS, 30.3), claims to have done just that. Concerning his Principle of Mood Constancy he says,

> . . . Phrastic validity plus satisfaction of the Principle provides necessary and sufficient conditions for the validity of mixed inferences. Those satisfying both phrastic validity and the Principle are valid . . . , those not satisfying both are invalid. . . . We are speaking here, of course, only of validity and invalidity in the "speaker-commitment" sense. But if Hare and MacKay are correct (and I think they are) this is the only sense applicable to mixed inferences (p. 103).

Clarke's Principle, (PC), is:

> The constituent elements of a mixed inference must be in the same mood for every occurrence within the inference (p. 101).

Unfortunately, it fails to live up to his claims.

In the first place, the mixed imperative *modus tollens* inference,

1. If you read the book, come to see me

A 2. Don't come to see me

3. Don't read the book,

is valid and yet it violates PC since the constituent element which occurs in the antecedent of premiss 1 recurs in the conclusion in a different mood.

Secondly, any mixed imperative instance of *hypothetical syllogism* will violate PC since, on the plausible assumption that imperatives cannot occur as antecedents of conditionals, all such inferences will have the form,

$$\begin{array}{ll} & 1.\ P \supset Q! \\ B & 2.\ Q \supset R! \\ \hline & 3.\ P \supset R! \end{array}$$[1]

in which the "middle term" 'Q' occurs in different moods in the consequent of the first premiss and in the antecedent of the second. Yet some mixed imperative instances of *hypothetical syllogism* are at least arguably valid.

---

[1] I am here, and throughout, following Clarke's notational convention of letting an exclamation mark after an expression indicate that the expression is in the imperative mood.

Thirdly, in view of the principle that adding premisses to an already valid argument cannot affect its validity, we can see that PC rules out many more valid mixed arguments. Consider the following recipe for constructing counter-examples: to any valid mixed imperative argument add a premiss which replicates one of the original constituent elements in a different mood. The resulting argument will be invalid according to PC since it will contain a constituent element which occurs in different moods, yet we shouldn't be able to transform a valid into an invalid argument by merely adding extra premisses.

So much for cases which show that PC does not provide a necessary condition for the validity of mixed inferences. Now for one which shows that it doesn't provide a sufficient condition either. If I am right in thinking that imperatives cannot occur as antecedents of conditionals, then any mixed imperative instance of *transposition* provides a counter-example to Clarke's sufficiency claim. This is because, while the inference,

1. $P \supset Q!$
C 2. $\sim Q! \supset \sim P$,

is both phrastically valid and does not violate PC, it cannot itself be valid since its conclusion is not well-formed.

I shall comment briefly on these four cases since each raises some interesting issues.

(*Case* 1) Inference A (the *modus tollens* case) is, of course, valid on Hare's original account (it is phrastically valid and contains enough imperative premisses to warrant an imperative conclusion) and on a logic of satisfaction (it would not be possible both to comply with (satisfy, terminate) the premiss commands and fail to comply with (satisfy, terminate) the conclusion command). It is also valid on a logic of "speaker-commitment". That is, a speaker who has commanded the premisses of inference A has thereby, or implicitly, commanded the conclusion. One thinks, *e.g.*, of Perry Mason telling his miscreant son, 'If you ever commit a capital crime, come to me for legal services' and 'Don't ever come to me for legal services' *as a way of telling him* not to commit a capital crime.

This means that on a logic of "speaker-commitment" a *modus tollens* inference can be valid while its corresponding *modus ponens* is invalid. Consider the *modus ponens* which corresponds to inference A,

1. If you read the book, come to see me
D 2. Read the book
3. Come to see me.

Here, as Professor Castaneda says, a student who went to see his teacher (solely) on the basis that he had issued the two premiss commands would be drawing the wrong conclusion. One wants to say that it's not enough just that the speaker issue the second premiss command: in order for the inference to go through, the second premiss command must be not only issued but complied with. But in the *modus tollens* case it *is* enough for the speaker just to have issued the two premiss commands.

(*Case* 2) When we consider counter-intuitive examples like,

1. If your feet are dirty, wash them
E 2. If you wash them, dry them off
3. If your feet are dirty, dry them off.

it might appear that we are better off without mixed imperative *hypothetical syllogism* inferences anyhow, and that PC does us a service by ruling them all out.

But in fact, (a) not all mixed imperative instances of *hypothetical syllogism* are as counter-intuitive as inference E, and (b) the counter-intuitive features of E can be accommodated (explained away) along the lines Hare uses[1] to accommodate the counter-intuitive features of imperative *simplification* arguments. That is,

(a) For a more plausible instance (in which the *hypothetical syllogism* is, as it were, embedded within (internal to) a predicate logic argument) consider Rescher's example:[2]

1. Make an abstract of every book you read
F 2. Read every book on the shelf
3. Make an abstract of every book on the shelf.

Paraphrasing this in Clarke's fashion, we get

1. (x) (Your reading x $\supset$ Your making an abstract of x!)
G 2. (x) (x being on the shelf $\supset$ Your reading x!)
3. (x) (x being on the shelf $\supset$ Your making an abstract of x!)

which is invalid according to PC since the "middle term" constituent element 'Your reading x' occurs in different moods in premisses 1 and 2. Yet inference F is not particularly counter-intuitive, at least not like E.

(b) What's troubling about the imperative *simplification argument*,

1. Put on your parachute and jump out
H 2. Jump out.

is quite similar to what is troubling about E. In both cases we want to say that, in light of the premisses, something important has been omitted from the conclusion command. Still, if Hare is right, the fact that something has been omitted from the conclusion is compatible with its being the case that nonetheless the speaker *is* (implicitly) committed to the conclusion command (having issued the premiss commands). And this is all that is required by the "speaker-commitment" conception of validity.

---

[1] R. M. Hare, 'Some Alleged Differences between Imperatives and Indicatives' (*Mind*, Vol. LXXVI (1967), p. 313).

[2] N. Rescher, *The Logic of Commands* (New York: Dover, 1966), p. 87.

(*Case* 3) PC's vulnerability to added premisses is shared by Hare's second rule which prohibits drawing an imperative conclusion unless there is at least one imperative among the premisses. The difference is that PC allows us (wrongly) to transform a valid into an invalid argument by adding an extra premiss whereas Hare's second rule allows us (wrongly) to transform the invalid argument,

I 1. You are going to jump in the lake
2. Go jump in the lake.

into a "valid" argument by adding the extra imperative premiss 'Go fly a kite'.

Unfortunately, the obvious modification which would eliminate this difficulty from Hare's second rule is not available for PC. All Hare needs to do is restrict application of his rule to just those sentential premisses whose phrastics are contained within the minimal set of phrastic premisses necessary for phrastic validity. The reason why this modification is available for Hare's rule but not for PC is that the phrastic of the added premiss in the counter-example to Hare's rule ('Your flying a kite') is different from the phrastic which comprises the minimal set necessary for phrastic validity ('Your jumping in the lake'). With PC, on the other hand, the phrastic of the added premiss is always going to be identical with some phrastic which *is* in the minimal necessary set. This is because the part in my directions for constructing counter-examples which says 'add a premiss which replicates one of the original constituent elements in a different mood' would read in Hare's terminology 'add a premiss which has the same phrastic but a different neustic from one of the original premisses'. *E.g.*, suppose, to the valid argument

1. If you will read the book, you will come to see me
J 2. You will read the book
4. You will come to see me,

we add the imperative premiss

3. Read the book.

This addition of an extra premiss will transform the original valid argument J into one which is (wrongly) designated invalid according to PC, since the resulting augmented set of premisses will contain occurrences of a constituent element in different moods in premisses 2 and 3. But since the phrastics of the added premiss 3 and the original premiss 2 are identical, it is hard to see how one of them could be relevant (necessary) and the other irrelevant to the *phrastic validity* of the argument. I suppose one might try to distinguish irrelevancy due to phrastic difference (*i.e.* irrelevant content) from irrelevancy due to redundancy or repetition, and invoke the latter sort for these cases, but so long as we stick strictly to questions of phrastic validity, while it is clear that one of the premisses 2 and 3 is phrastically redundant, it is not clear which one is, nor that it makes any sense to ask.

(*Case* 4) I take it to be a fact that imperatives cannot occur as antecedents of conditionals. Using Clarke's notation, I claim that the formula '$P! \supset Q$' is not well formed (or, is deviant). Now it might be objected that this is only apparently the case, due to our attempting to read '$P! \supset Q$' as, *e.g.*, 'If come to see me, then you will read the book'. But (the objection runs) we need not require that '$P! \supset Q$' be mechanically transcribed in the 'If P then Q' fashion. We can have recourse to the alternative reading 'P! only if Q' (*i.e.*, 'Come to see me only if you read the book'). Introductory logic texts assure us that 'P only if Q' is just a stylistic variant of 'If P then Q', and how else can we understand 'Come to see me just in case you read the book' except as 'Come to see me if you read the book' and 'Come to see me only if you read the book'?

I think that most of these observations are correct, but that the conclusion drawn is wrong. What these remarks show is not that imperatives can occur as antecedents of conditionals, but that the locution '... only if...' is not always just a stylistic variant of the locution 'If...then...'. The command 'P! just in case Q' does break down into the pair of conditionals 'P! if Q' and 'P! only if Q', but in both of these the imperative occurs in the consequent of the conditional. This is because the command 'Come to see me only if you read the book' has to be understood in the transposed form 'If you don't read the book, don't come to see me'.[1] That is, 'Come to see me just in case you read the book' breaks down into the conjunction of 'If you read the book, come to see me' and 'If you don't read the book, don't come to see me'. This wouldn't be worth remarking if the standard transpositional equivalence between '$\sim Q \supset \sim P$' and '$P \supset Q$' held for conditional imperatives; but it doesn't, and so it is.

Now, as I pointed out above, any mixed imperative instance of *transposition* (like inference C) is both phrastically valid and does not violate PC, yet it cannot be valid since it involves a deviant conclusion. At least the sufficiency half of Clarke's claim about PC needs to be modified to say something like this: Given that a mixed inference is phrastically valid, and does not violate PC, and does not lead to a deviant conclusion, then it is valid in the "speaker-commitment" sense. Ideally, of course, an adequate set of formation and transformation rules would not even allow the possibility that well-formed premisses and otherwise valid inference patterns might result in a deviant conclusion, but in default of these, some such qualification needs to be made to the statement of PC.

---

[1] *Ibid.*, p. 33, '. . . The command 'X do A only if B occurs!' $\cong$ [X! refraining from doing A/B does not occur].'

Thus I conclude that Clarke's Principle of Mood Constancy, together with phrastic validity, does not provide necessary and sufficient conditions for the validity (in the "speaker-commitment" sense) of mixed inferences.

Oberlin College

"JONESESE" AND SUBSTITUTIVITY

By A. C. GENOVA

PROFESSOR ZEMACH'S article 'Reference and Belief' (ANALYSIS 30: pp. 11–15) represents another attempt to defend the principle of substitutivity (sometimes erroneously labelled 'Leibniz's Law') and account for the apparent failure of extensionality in the context of *oratio obliqua* constructions. I want to argue that Zemach's recommendations are fundamentally incoherent and generate more problems than they purport to resolve.

Zemach wants to replace Quine's distinction between referentially opaque and transparent contexts with a simple substitution principle which (1) takes account of the distinction between 'proper-use' (the use of a term, x, to refer to whatever the present user of the term himself would use it to refer to) and 'other-use' (the use of the same term to refer to whatever somebody else—typically somebody who is mentioned in the sentence containing x—would use it to refer to); and (2) permits unlimited substitution, *salva veritate*, of all co-referential terms in all linguistic contexts just so long as the substitutive terms are 'proper-used'. Accordingly, Zemach contends that if

(1) x is a Greek=x is a Hellene.

is true in our language, then if

(2) Jones believes that every Greek is a Greek.

is true, it necessarily follows that

(3) Jones believes that every Greek is a Hellene.

is also true if 'Hellene' is proper-used in (3)—even if Jones' vocabulary lacks 'Hellene' or he is unaware of its established usage. Otherwise, if 'Hellene' is other-used in (3), then (3) is not entailed. Zemach's point is that if 'Hellene' is proper-used in (3), then we are merely attributing to Jones a particular belief about certain entities which *we* (*sc.*, the utterer of these sentences) refer to by 'Greek' or 'Hellene' without thereby claiming that this is also Jones' mode of referring to these entities. I take it that Zemach would also argue that if 'Hellene' is other-used in (3), then the argument is invalid because 'Hellene' would coincide with Jones' mode of referring, and then we could never be sure (without additional information) that it would correlate with our standard use of 'Hellene'. After all, as Zemach suggests, Jones might be using 'Hellene' to refer to the property of being black (in the language of "Jonesese").

This approach is set forth as an amendment to Professor Sellars' view (ANALYSIS 15: pp. 117–120) which also supported the entailment of (3) in opposition to Professor Putnam (ANALYSIS 14: pp. 114–122) on the narrower ground that those who disallow the inference are failing to heed the use-mention distinction and take inadequate account of synonymy relations of terms in our language. Zemach's amended version does not rest on the use-mention distinction, but rather on the possibility of homonymous uses of the same expression by different users of the language—a possibility which supposedly becomes manifest when we relativize the use of referring expressions to individual speakers; and moreover, Zemach's version extends over all cases of sameness of reference, not merely synonymous terms.

Now it is precisely this amendment, I think, that gets Zemach into trouble. Sellars' position is much more conservative because Sellars' comments apparently only apply to the substitutivity of logical equivalences or expressions which have the same meaning, not to all co-referential terms; and further, in resting his case on the use-mention distinction, he would presumably have to agree with Quine that not only are quotational contexts referentially opaque, but also that if there *were* other contexts in which substantial expressions were not used to refer at all, then these contexts could be opaque and therefore not admit of unlimited substitution of non-synonymous extensional equivalents. But Zemach's principle allows unlimited substitution of all co-referential terms in all contexts, and unlike Sellars', his position implies that if Jones uses a substantival expression to refer to a different entity (or kind of entity) than the correct users of English use it to refer to, then Jones and the correct users of English are respectively speaking in different languages. He argues that if Jones believes that 'Hellene' means 'black', then 'Hellene' is '. . . indeed *used* in (3), but it refers to whatever "Hellene" refers to *in Jonesese*, *i.e.*, to the property of being black'. Again, he says 'The only question that we still have to ask will be, in what (or, in whose) language is the problematic term, x, being used'. Finally, in the context of other examples, he implies that if a definite description like 'the murderer' or 'the richest man in town' is used by two speakers of English to refer to different, particular entities, then these different instances of the use of this singular expression correspond to nothing but homonyms; and similarly with 'Nixon' if the latter is used by Tibetans to refer to Chairman Mao while we use it to refer to the 36th president of the United States (pp. 13–14).

Now what does Zemach's thesis amount to? It seems to me that the primary question which confronts us when we examine *oratio obliqua* constructions is whether or not we are to count sequences like (1), (2) and (3) as instances of valid *arguments*, *i.e.*, does the conjunction of the statements expressed by sentences (1) and (2) entail the statement expressed by sentence (3)? Now surely the notion of evaluating arguments expressed in a language presupposes the fact of a shared, common language within which these arguments are expressed. We assume that the following law is true: $F_a \equiv (\exists x)\ [F_x . (x=a)]$. To deny this seems self-contradictory and certainly counter-intuitive because its denial would imply that there could be an x, *e.g.*, b, which has the property F and is equal to a, and yet $\sim F_a$ is true. This law, in turn, assumes the truth of Leibniz's principle of the indiscernibility of identicals: $(x)(y)\ [(x=y) \supset (F)(F_x \equiv F_y)]$. This latter principle becomes linguistically reformulated as 'the principle of the substitutivity of identicals' which allows unlimited substitution, *salva veritate*, of extensionally equivalent *expressions* in the relevant containing *sentences*. Thus, it is this substitutivity principle (not Leibniz's Law) which becomes doubtful in the context of indirect discourse.

The *problem* is that although the denial of this substitutivity principle also seems counter-intuitive, yet, in certain contexts of ordinary discourse, we are clearly unwilling to permit the problematic substitution because it seems to do violence to what we want to say. For example, if the following two assertions are true,

(4) Mary = the prostitute on the third floor.

(5) Jones believes that Mary is a virtuous woman.

are we to say that they entail

(6) Jones believes that the prostitute on the third floor is a virtuous woman.

and leave it at that? Well, we are inclined to say yes in one sense, no in another. My point is that our ambivalence here should not embarrass us. Surely the *fact* that sentences like (3) and (6) admit of a dual interpretation is a legitimate fact about our natural language. The philosophical problem associated with them concerns how to account for this fact in some general way, *i.e.*, how to expose the logical grammar of such locutions while neither doing violence to our intuitive adherence to the substitutivity principle, nor over-abstracting from the context of the vernacular such that we find ourselves forced to say what we do not want to say.

I submit that the philosophical issue hat stake here as nothing to do with how the mentioned believers in sentences like (3) and (6) may have actually employed the problematic referring expressions in datable, individualized acts of referring. If I understand Zemach correctly, he is saying that (3) and (6) are false only when the respective, problematic referring expressions are other-used, *i.e.*, are related to the proper-use of these expressions by the relation of homonymy. Leaving aside for the moment Zemach's notion of what constitutes homonymy in a language, let us examine his revised substitutivity principle more closely. Now in respect to any such problematic term, x, it would seem that x might be related to the so-called "Jonesese" language in three possible ways: (a) Jones may have no use for x at all because it is not an item in his vocabulary; (b) Jones uses the term to refer to something categorically different from the referent of its proper-use; or (c) Jones uses it to refer to the same kind of individual as we do, but to a different individual.

Now one difficulty might be as follows. Suppose (as Zemach allows) that Jones is not at all familiar with the term 'Hellene'. Then I take it that 'Hellene' could not be other-used in (3) because it could not be used at all by the mentioned believer in (3)—there is no use for the term in "Jonesese". Then, either 'Hellene' in (3) is not used to refer at all, or it is proper-used. The first alternative would require us to say that the belief context here exhibits Quine's referential opacity—an interpretation that Zemach wants to do without. But the second alternative would imply that the argument is invariably valid, in spite of the fact that Zemach seems to want to allow for a sense in which it can always be invalid. Thus, if the problematic term does not occur in "Jonesese", then it would seem that either the belief context is opaque in (3), or (3) is invariably true.

Zemach apparently applies his principle indiscriminately not only to cases of synonymous and non-synonymous co-referential terms, but also to singular terms of all shades, *e.g.*, names ('Jones'), general terms ('Greeks'), indefinite descriptions ('a Greek') and definite descriptions ('the richest man in town'), even though these terms involve different kinds of referents and different criteria for correct usage. Limitations of space do not permit an adequate discussion of how these differences affect the problem of substitutivity in oblique contexts, but the most pregnant example is provided by the case of non-synonymous, co-referential, definite descriptions. For when the problematic expression is a definite description, it seems quite possible that the mentioned believer in sentences like (6) can *mean* the same thing that everyone else means by the expression, and *use* the expression in the same way that everyone else uses it, and nevertheless, the problem of substitutivity still remains.

Zemach wants to say that when the problematic inference is invalid (that is, when (6) is treated as false), the invalidity is due either to the misapplication of standard criteria of reference or to the correct application of deviant criteria of reference. But is it either of these if Jones simply did not know or realize that the prostitute on the third floor was identical with Mary? Surely Jones may mean the same thing we all mean by 'the prostitute on the third floor' and may use this description to refer to the very same entity which the speaker of (6) may use it to refer to, and yet, may not have recognized (for all sorts of contingent reasons like the fact that Mary was good at disguising herself when she performed as the prostitute) that the relevant entity also admitted of the designation 'Mary' or vice versa. It isn't that Jones was unfamiliar with the description; it isn't that Jones radically misused it to refer to the property of being black (deviant criteria); it isn't that Jones used it to refer to a different entity (misapplication); it is just that Jones, while perfectly capable of using the description, either does not have the *occasion* to apply it, or applies it correctly to the entity who is Mary but doesn't recognize it *as* Mary. Thus, it seems that this constitutes a counter-example to Zemach's principle, because 'the prostitute on the third floor' could be proper-used in (6) and there still are good grounds for treating (6) as false.

Zemach would probably maintain that this would still be a case of misapplication because, in not recognizing the prostitute as Mary, Jones misapplied the criteria of personal identity, *viz*., the criteria for recognizing the spatio-temporal continuity of Mary and the prostitute. But this seems too strong. If it were suddenly discovered that one of our close friends was identical with the unsavoury panhandler we passed every night on the way home, would it make sense to say that we misapplied the standard criteria of identity for our friend? But even if we allowed all this, to suggest that Jones' use of the expression 'the prostitute on the third floor' was an occurrence in "Jonesese", while the utterer's use of it in (6) was an occurrence in English, seems patently absurd.

There are other difficulties, or at least questions, which pertain to Zemach's substitutivity principle. In saying that the principle is to be applied 'in all (including oblique) contexts' (p. 13), I assume that Zemach would employ the principle not only to belief contexts, but also to those oblique constructions which include other verbs of propositional attitude ('knows', 'desires', 'thinks', 'imagines', *etc.*) as well as other presumably non-extensional contexts involving non-psychological prefixes as exemplified in alethic or causal modalities. But it is not at all clear how the principle would fit these cases. For example, how would the other-use/proper-use distinction apply to the well known chestnut concerning whether 'It is possible that the morning star is not the morning star' must be true if the conjunction of 'Venus is identical with the morning star' and 'It is possible that Venus is not the morning star' is true?

Additional questions arise concerning the other apparent failures of extensionality which occur in these contexts—I mean failures other than those involving the substitutivity of co-referential terms. For example, there are problems concerning the permissibility of existential generalization in respect to substantival expressions occurring in the embedded sentences exhibited in the oblique containing sentences. Does 'Jones believes that Mary is a virtuous woman' entail 'Mary exists'? What we say here may well turn on whether we view the belief context as transparent or opaque, *i.e.*, whether or not 'Mary' has a referential use, but I cannot see how the issue can depend on whether 'Mary' is other-used or proper-used.

Finally, what about the fact that these non-extensional contexts also prohibit the replacement of the embedded sentences with sentences of the same truth value? If 'Mary is a virtuous woman' and '$2+2=4$' are both true, then clearly, 'Jones believes that Mary is a virtuous woman' does not entail 'Jones believes that $2+2=4$'. But why not? Would it make sense to say that the substitution fails because '$2+2=4$' is other-used in the latter assertion? I am not suggesting that Zemach would say these things, but merely that his substitutivity principle would fall far short of solving most of the difficulties connected with oblique contexts, while Quine, rightly or wrongly, can treat most of the problems with his single principle of referential opacity.

In respect to the general problem of substitutivity in oblique contexts, I have been arguing that it cannot be solved by relativizing the use of referring expressions to individualized modes of referring, and then making the validity of the problematic substitution depend on whether the tokens of the referring expressions are expressed in English or some private language like "Jonesese". The genuine problem concerns how we are to reconcile what we consider to be sound principles of deductive inference with the apparent non-extensionality exhibited by the linguistic phenomenon of indirect discourse. It also concerns, I suppose, the need for an adequate logic of modal prefixes as well as a clear understanding of the relation between our canonical notation and our natural language.

The logical regimentation of our ordinary ways of talking is not a way of *improving* on the latter. It is not as if our native tongue were infected with vagueness and ambiguity, and symbolic translation were the means of improving communication. This would be like putting the logical cart before the linguistic horse. For example, the translation of natural utterances into quantificational form is a way of revealing the already embedded logical grammar of ordinary discourse, not a way of revising it. In this sense, the symbolic notation is a kind of commentary on everyday discourse—and it is from the latter that we should take our cues. The point of logical analysis is to uncover the general logical pattern which already constitutes part of our overall understanding of how to speak and follow the norms of usage. Thus, logical analysis presupposes the background of a public language, a shared usage of expressions *in* a language; and accordingly, questions about validity, substitutivity, reference, synonymy or homonymy become quite irrelevant if someone decides to abstract himself from this shared linguistic frame of reference and speak in "Jonesese".

In recent discussions of the problem of substitutivity in oblique contexts, there have appeared two general lines of solution. The first approach might be labelled a 'logicist' or 'microscopic' approach and is perhaps best exemplified by Quine (*Word and Object*, M.I.T. Press, 1960, pp. 141–51). The second approach might be called a 'contextualist' or 'macroscopic' approach and is well represented by Leonard Linsky (*On Referring*, Humanities Press, 1967, pp. 114–15). Quine more or less abstracts from the actual contexts of speech acts and focuses on the inner logic of the "propositions" or "statements" expressed by the sentences of our natural language; while Linsky examines pieces of linguistic behaviour in the actual context of the utterances, thereby taking account of the intentions of the speaker, the audience, the psychological circumstances, and the external linguistic conventions which govern the use of our ordinary discourse.

Typically, Quine treats sentences like (6) above by construing the scope of the peccant verb as covering an intensional abstraction (an embedded proposition) within which substantival expressions do not function as referring expressions at all (they have non-referential position), and hence, the overall containing sentence (the belief context) is referentially opaque (and therefore non-extensional). Accordingly, the apparent failure of substitutivity in these cases does not constitute an exception to the substitutivity principle, but merely a case of failure of reference, *i.e.*, non-referential position within an opaque context. But this interpretation of (6) would not be a rigid one for Quine because he clearly insists that we cannot take the belief construction as *invariably* opaque or transparent. We always have the option of reading (6) so that 'Mary' achieves referential position (and thereby permits substitution), so long as we recognize that this requires an appropriate narrowing of the opaque scope and a different reading of (6). Thus, if 'Mary' in (6) occupies an opaque position, then (6) means something like 'Jones is related, by the relation believing, to the proposition *that Mary is a virtuous woman*'; while if 'Mary' in (6) is construed as occupying a transparent position, then (6) means something like 'Jones and Mary are related by the relation of believing-to-be-a-virtuous-woman', *i.e.*, Jones here is related to another individual, not a proposition.

In contrast to Quine, Linsky interprets instances of referring expressions as always referring to what we ordinarily use them to refer to in their appropriate linguistic contexts. He reminds us how the *oratio recta* and *oratio obliqua* modes actually function in our natural language. The utility and power of the device of indirect discourse rests precisely in the fact that it enables us to communicate the essential content without the need to duplicate the exact language we are reporting, *i.e.*, in this mode, we are not responsible for reproducing the very words in which another person (the mentioned believer) would express his own propositional attitudes. Moreover, the mode of indirect discourse is itself regulated by linguistic coventions tied to the circumstances of utterance such that no one *expects* us to say anything other than what we are saying (assuming that our audience also understands the relevant conventions). Linsky argues that there are no fixed limits to the use of this mode, but that the fundamental principle is that one must not mislead one's audience; and this, in turn, is dependent upon the complex facts of the situation in which one speaks. So there are good reasons for the so-called failures of substitutivity, and their successful elimination would also erase the essential device of indirect discourse. Thus, it is only the overly zealous logician who would be puzzled if one said that 'Oedipus wanted to marry his mother'. The audience, say, of students who had read the play, would not be bothered at all.

In conclusion, it is obvious that neither of these approaches depend upon distinctions between proper-use and other-use, or English and "Jonesese". Zemach's approach won't do because if the mentioned believer is speaking in our language, then the problem of substitutivity remains even if he uses the problematic referring expressions just as the rest of us do; and if he is not speaking in our language, but in "Jonesese" instead, then it isn't a problem in our language at all.

Wichita State University

BELIEF STATEMENTS AND THEIR LOGIC

By F. H. GEORGE

OVER the years a number of articles have appeared referring to belief statements, beliefs and their relation to logic. One of the first articles to appear on this subject was by Church (ANALYSIS, 10.5) who took two key sentences:

Seneca said that man is rational ... (A)
Columbus believed the world to be round ... (B)

Church stated that these were both statements about abstract entities called propositions. He then went on to argue that the "translation test" of statements such as (A) and (B) from English into any other language made it clear that one could believe a statement in one language and not believe the equivalent statement in another. This is obviously so since one can hardly be said to believe in a statement that one does not understand. The nub of this argument, as of other similar arguments about belief statements which follow from Putnam, Sellars and, more recently, Zemach, (ANALYSIS, 14.5, 15.5 and 30.1 respectively) has been the intensional isomorphism of Carnap (*Meaning and Necessity*) although other recent explications have relied on a distinction between so-called opaque and transparent beliefs.[1]

Intensional isomorphism analyses sentence B as follows:

There is a sentence B in the semantic system S such that (1) *B* is intensionally isomorphic to 'the world is round' and (2) *Columbus was disposed to an affirmative response to B.* This sort of explanation, originally suggested by Carnap, has been modified from time to time by various people and one could especially point to Putnam's revision here since he recognised that it was possible to argue that:

> It is possible to doubt that whoever believes that all Greeks are Greeks believes that all Greeks are Hellenes. ... (C)

He went on to argue further that intensional isomorphism in Carnap's terms requires the extra condition that the two forms ('all Greeks are Greeks' and 'all Greeks are Hellenes' in (C)) referred to were logically equivalent.

The difficulty with the whole of this discussion is the very fact that we talk about the *logic* of belief statements. This suggests that the matter of the intensional relations involved is independent of the users of the statements themselves which manifestly they are not. It is clear from

---

[1] Sellars, W. 'Some Problems about Belief', *Philosophical Logic*, (J. W. Davis, J. J. Hockney and W. K. Wilson, Eds.) Reidel.

Carnap's original statement of intensional isomorphism that he is dealing with a pragmatic matter. What makes it clear is section (2) of his definition which specifies that a particular person, Columbus, was disposed to make an affirmative response to (B). In other words, the amount of information that Columbus had acquired at that time determined the belief he would maintain with respect to some statement, or perhaps the proposition that the statement referred to.

Thus it is that Zemach's more recent discussion of the point makes it clear that "proper use" and equivalent references are the things that really matter if two beliefs are thought to be intensionally isomorphic to each other. In other words, to use Sellars' terminology:

Bfa
a=b
Bfb

where Bfa is some such statement 'Jones believes that fa' and Bfb is some such statement as 'Jones believes that fb'
and we shall conclude

Bf(a=b)

(Jones believes that a=b).

In other words, there is absolutely no problem to solve if we accept the two critical conditions of a pragmatic kind which apply in such situations, *i.e.* as Zemach puts it, that the references of the two terms or phrases are the same and that regardless of the way the phrases are translated or expressed in statements, that the person who utters them believes the references to be the same.

If the person uttering the statement merely believes two terms to have the same reference when they do not, then he is mistaken and the belief statements are not equivalent. If on the other hand, the two terms or phrases refer to the same thing and they are believed not to by the person making the statements, they still do not amount to equivalent belief statements because the operative word is 'belief' which makes it clear the context of the logic is a pragmatic one. Thus it is that the necessary and sufficient condition for equivalent belief statements is that the two terms or phrases have to have the same reference and the people uttering the statements have to believe that they have the same reference. This is all that can be said about the logic of belief statements in this regard and makes it clear that Carnap's original version of intensional isomorphism is essentially right as it stood.

Brunel University

SOME DOUBTS ABOUT ILLOCUTIONARY NEGATION

By Richard T. Garner

In 'Transformations of Illocutionary Acts' (Analysis, 30 (1969), pp. 56–59) Aaron Sloman discusses the concept of illocutionary negation found in John Searle's new book *Speech Acts* (Cambridge, 1969). What seems to be the same concept appears in R. M. Hare's 'Meaning and Speech Acts' (*The Philosophical Review*, 79 (1970), 3–24) under the name of external (to be contrasted with internal) negation. This paper will be a discussion of some substantive and methodological issues that arise in connection with that concept, though not of the use made of it by Sloman and Hare in dealing with problems about meaning.

I

In the final chapter of *How to do things with Words* (Harvard, 1962) J. L. Austin constructed lists of verbs which 'make explicit . . . the illocutionary force of an utterance, or what illocutionary act it is that we are performing in issuing an utterance' (pp. 148–49). It is in this tradition that Searle distinguishes first between the illocutionary act and its propositional content, and then between:

> two (not necessarily separate) elements in the syntactical structure of the sentence, which we might call the propositional indicator and the illocutionary force indicator. The illocutionary force indicator shows how the proposition is to be taken, or to put it another way, what illocutionary force the utterance is to have (p. 30).

Sometimes we employ explicit syntactical devices to indicate the intended force of our utterances, and sometimes 'the context will make it clear what the illocutionary force of the utterance is, without it being necessary to invoke the appropriate explicit force indicator' (p. 30). Having made this point, Searle introduces some symbolism:

> The general form of (very many kinds of) illocutionary act is
> $F(p)$
> where the variable 'F' takes illocutionary force indicating devices as values and 'P' takes expressions for propositions (p. 31).

For example, assertions are symbolized '$\vdash (p)$', promises '$Pr(p)$' and yes-no questions '$?(p)$'. One advantage claimed for the distinction between the two kinds of indicators is that it enables 'us to account for and represent the generally overlooked distinction between illocutionary negation and propositional negation' (p. 32). For example, 'I promise to come' is said to have an illocutionary negation:

(1) I do not promise to come. $\sim Pr(p)$

and a propositional negation:

(2) I promise not to come. $Pr(\sim p)$

In 'Meaning and Speech Acts' Hare subscribes to a very similar view about negations:

> The internal negation of 'I promise to pay you before the end of the tax year' is 'I promise not to pay you before the end of the tax year'. The external negation of the same promise is 'I do not promise to pay you before the end of the tax year' (p. 12).

After adding that 'nearly all speech acts, including assertions, can be negated in these two ways' Hare contrasts the internal and external negations of 'The cat is on the mat'. The external negation is:

(3) I don't say that the cat is on the mat. $\sim\vdash(p)$

while the internal negation is

(4) The cat is not on the mat. $\vdash(\sim p)$

In recent years linguistic philosophy and philosophical linguistics have combined to shake our faith in information gleaned from an unsupplemented examination of the surface structure of a sentence. There are many ways to go beyond such a method. Searle, for example, begins with a set of grammatically related sentences and asks: 'What shall we say the speaker is doing when he utters one of these?' (p. 22). In answering this question he is led to distinguish the two components mentioned above and to propose his symbolism for representing them. With this I take no issue. However, I shall try to show that while some sentences admit 'not' in (at least) two semantically relevant positions, and while propositional negation functions in a *relatively* unproblematic way and can be correlated with one way of negating sentences, there is less reason to suppose that the form of *sentence* negation exhibited in (1) and (3) is correlated with a kind of negation at the illocutionary act level.

II

Sloman seems to accept the distinction between illocutionary and propositional negation, though he is not totally satisfied with Searle's suggestion that one who utters (1) is *refusing* to make a promise. He offers his own view of illocutionary negation in two passages:

> (A) The effect of 'not', or 'I refuse to' on an illocutionary act indicator seems to be simply to produce a non-committal utterance (pp. 57–58).
>
> (B) If F is a sign indicating performance of a certain speech act, then the effect of 'not' on that sign is to produce a new sign which indicates (but does not state) that the original speech act is under consideration but the speaker is not yet prepared to accept the commitments involved in performing it (p. 58).

Recall that Searle first explained illocutionary force indicators and proposition indicators, and only later introduced symbolism to represent them. Though he speaks (somewhat misleadingly, I think) of the 'negation of an illocutionary force indicating device', nothing he says warrants Sloman's talk in (A) about 'the effect of "not", or "I refuse to" on an illocutionary act (*sic*) indicator'. If the illocutionary force indicator shows 'what illocutionary force the utterance is to have' (Searle, p. 30), then someone using a sentence of the form 'I refuse to X' as an explicit performative is employing one and only one illocutionary force indicator, *viz.*, the explicit 'I refuse to'. In particular, one who says 'I refuse to promise to come' is not using an illocutionary force indicator of the sort explicitly present in (2). Hence, contrary to what is assumed in (A), there is no illocutionary force indicator available in 'I refuse to promise to come' for 'I refuse to' to have an effect on.

Searle points out, and quite correctly, that 'illocutionary force indicating devices in English include at least: word order, stress, intonation contour, punctuation, the mood of the verb, and the so-called performative verbs' (p. 30). It is incorrect, then, for Sloman to speak in (A) of the effect of 'not' on such *indicators*, and in (B) of the effect of 'not' on 'a sign indicating the performance of a certain speech act'. It is quite likely that a confusion between the symbolism and what it was introduced to symbolize has led Sloman to say things about the latter that only make sense when said about the former. If so, it would be advisable to reformulate (B) in the following way:

> If F is a sign representing a device used to indicate performance of a certain speech act, then the effect of 'not' on that sign is to produce a new sign which represents a device used to indicate (but not used to state) . . . *etc.*

As soon as this adjustment is made it becomes clear that certain priorities have been displaced, and that in fact Searle's symbolism has become detached from the distinction it was originally introduced to represent. When first introduced, it was merely to stand for a set of diverse ways to indicate illocutionary force, and in its capacity as an abbreviation required no interpretation. In Sloman's hands, however, the symbolism seems to have begun to influence the sort of analysis it once served in a mere notational capacity. He seems to be saying: 'these permutations are possible, so let us ask what they mean and if they don't mean anything in English, let us ask why not' rather than 'let us see what sorts of illocutionary acts people who utter externally negated explicit performative sentences can perform, and then see whether we need a complex sign to represent a certain kind of illocutionary force indicator'. I want to suggest that the second question re-establishes the proper priorities, and that the rather drastic step of elevating illocutionary negation from sentences to speech acts should only be taken if an adequate treatment of language necessitates such a move. I do not believe that it does. In fact, I believe that one who begins as I have suggested will find himself inclined to interpret Searle's remark that 'illocutionary negations in general change the character of the illocutionary act' (p. 32) as saying only that an explicit performative sentence and its external negation are generally used in the performance of different kinds of speech acts. One may then wonder whether anything remains to be said that can be expressed only, or even best, by speaking about negations, or more generally about transformations, of speech acts, rather than of sentences.

III

We have seen Hare claiming that 'nearly all speech acts . . . can be negated in these two ways', *i.e.*, externally and internally. But if the very idea of external negation *of a speech act* is suspect, the doctrine that nearly all can be negated in this way is even more so. To be sure, just about any sentence can easily be cast into other sentences which make explicit one or another of its potential illocutionary forces. And these sentences can, in turn, be symbolized in the notation suggested by Searle. Finally we may, if we wish, pair each explicit performative sentence with two other sentences, one of which is its external, and the other its internal, negation. Such an operation is analogous to inserting a '$\sim$' before either '*F*' or '*p*' of a formula in Searle's notation. But the fact that we can generate such structures does not guarantee that the sentences so formed will have a use, nor, granting that some will, does it provide any indication of what that use might be.

Consider performing external negation on the following sentences:

I give and bequeath my watch to my eldest son.
I declare this bridge open.
I bid you welcome.

I warn you—that line carries high voltage.
I remind you that the deadline is October 31.

With a little effort one could doubtless construct cases in which the external negations of the first three sentences have a non-constative use that would not be considered hopelessly artificial or deviant. The external negations of the other two are, for obvious reasons, best ignored. But what reason could there be (other than the superficial grammatical appearance of 'not') for describing the speech acts that might be performed by uttering the external negation of the first three sentences as the negations of speech acts usually performed by uttering the original non-negated counterparts?

By any criterion of illocutionary acthood, refusing seems to qualify as an illocutionary act. And certainly we would, at least sometimes, perform acts of refusal by uttering externally negated sentences. I may refuse to run a mile, pay my debts, get married, make a promise, or impart some information; and whenever I do so in words, my illocutionary act belongs to the class of refusals—just as my acts of promising to run a mile, pay my debts, get married, make a promise, or impart some information, belong to the class of promises. Refusing to promise (whatever form of words is adopted) is no more a modification or transformation of the act of promising than refusing to run is a modification or transformation of the act of running. It is this very important point that it seems to me those who want to speak about negations of speech acts have missed.

IV

The belief that external negation is a feature of a speech act rather than of sentences has unfortunate consequences. For example, it leads naturally to speculation about the effects of other logical operations on speech acts. Thus Sloman asks if there 'could not be a use of "if" which qualifies a speech act by making it provisional' (p. 68). He appears to take the affirmative answer to this question rather seriously, for he presents, rather sympathetically, an analysis according to which:

> utterances of the structure 'If F (p) then G(q)' express provisional commitment to performing G on q, pending the performance of F on (p). Utterances of the form 'F(p) or G(q)' would express a commitment to performing (eventually) one or other or both of the two acts though neither is performed as yet. The question mark, in utterances of the form 'F(p)?' instead of expressing some new and completely unrelated kind of speech act, would merely express indecision concerning whether to perform F on p together with an attempt to get advice or help in resolving the indecision (p. 58).

Presumably anyone uttering the sentence:

(5) If it rains then it will get colder.

is expressing a provisional commitment to "performing assertion" on the proposition that it will get (is getting?) colder "pending the performance of" assertion on the proposition that it rains. But what can this possibly mean? Fortunately Sloman provides us with an explication of what one who utters (5) is doing; but unfortunately his explication differs from the one just constructed from his generalized formula. He says that one uttering (5) would be 'provisionally asserting' that it will get colder. Perhaps further explanation would make this more palatable, but for now the truth of the matter seems to be simply that one who utters (5) is, then and there, asserting that if it rains it will get colder, and provisionally asserting nothing at all.

Assuming (5) to be of the form 'If *F*(*p*) then *G*(*q*)', it is very difficult to know how to determine what other sentences count as having that form—especially if '*F*' and '*G*' should be replaced by symbols standing for non-constative illocutionary force indicating devices. Sloman remarks that the result of putting 'commendation or endorsement in place of the speech acts F and G in the above schemata . . . seem to correspond moderately well with some (though not all) actual uses of the words and constructions in question' (p. 58). This suggests that we might be able, given certain obvious abbreviations, to make sense out of such formulae as: 'If *Comm*(*p*) then *End*(*q*)','If *End*(*p*) then *Comm*(*q*)', and 'If *End*(*p*) then *End*(*q*)'. But I find the results of the suggested replacement so strange that I do not know how to go on to try to determine whether or not they correspond with anything. Sloman just doesn't say enough to allow us to determine whether or not anyone might utter a sentence, and in so doing perform some speech act that would be best representable by any one of the above formulae.

Sloman's treatment of 'or' in the passage quoted above does not clear any of these questions up. Since (5) is Sloman's own illustration of a sentence of the hypothetical form, it is natural to suppose that the disjunctive form would be illustrated by something like:

(6) The music has stopped or I'm losing my hearing.

But if we examine what is said about 'utterances of the form "F(p) or G(q)" '

(7) I will promise to cut the grass or I will promise to wash the dishes.

seems a more likely candidate. Waiving points about various senses of 'or', (7) at least does 'express a commitment to performing (eventually) one or other or both of the two acts though neither is performed as yet'.

It may be, however, that the form '*F*(*p*)?' (which is not even well-formed in Searle's symbolism) is the most baffling of the three discussed in the quoted passage. Sloman's treatment of it betrays once more a confusion between the illocutionary force indicators used in speech and the symbols introduced to represent them. He says that the question mark in 'utterances of the form "F(p)?" ' expresses 'indecision concerning whether to perform F on p together with an attempt to get advice or help in resolving the indecision'. Yet punctuation is but one method of indicating illocutionary force, and it is hardly relevant when spoken discourse is in question. The question mark occurs in the *formula*, and accordingly it can only be supposed to represent a device used to convey an intended illocutionary force. But on the given understanding of illocutionary force indicators so does any value of '*F*'. Consequently, the formula '*F*(*p*)?' is incoherent. One may try to say that it can represent a question (the only plausible alternative) that one might ask by uttering the words 'Should I promise to marry her?', but then I fail to see why it is needed, for we already have the form proposed by Searle for yes-no question: '?(*p*)'.

The moral to be drawn from this is that when we begin with symbols, or allow them to take over at some point, we become open to confusions and distortions that might easily have been avoided. Only if we permit ordinary language really to have the "first word" and begin by examining hosts of sentences like:

> If you don't have a gas mask I advise you not to visit Chicago.
> Go to the store or suffer a horrible death.
> If you love me keep my commandments.
> Put up or shut up.
> If you don't tell on me I promise never to do it again.

do we have a hope of understanding what is going on when we find complex sentences whose components seem to contain illocutionary force indicators. Only when we have begun to understand the workings of such sentences should we raise the question of whether or not a symbolism is needed, and if so, of what form it should take. I doubt that it will ever be necessary to construct "pseudo-truth-functional formulae" for this purpose. Apart from exceptions of the sort mentioned by Searle on pp. 70–71, anyone who accepts the distinction between illocutionary force indicators and propositional indicators, and then employs symbolism of the type introduced by Searle, will have no need for anything more complex than formulae which contain one and only one un-negated symbol serving as the value of 'F' in a formula of the form '$F(\phi)$' (where 'ϕ' includes, but is not limited to 'p', and contains no illocutionary force indicators).

The Ohio State University

8 JUL 1971

NOTES

The Analysis Committee consists of: Chairman, P. T. Geach; Secretary, Miss M. A. Boden; J. H. Benson, Alec Fisher, A. Harrison, R. F. Holland, S. Körner, A. C. MacIntyre, A. M. MacIver, B. Mayo, R. G. Swinburne, A. R. White. This committee is responsible for appointing and advising the editor and for the general policy of the paper.

Subscriptions. The subscription to Analysis, Vol. 31, is 24s. (£1.20) in Great Britain and 28s. (£1.40) abroad (U.S.A. $4.00), post free. Each volume comprises six numbers, five of 32 pages and one of 48 pages, appearing within the academic year—in October, December, January, March, April and June. Orders should be sent to Basil Blackwell, 49 Broad Street, Oxford, or placed with any bookseller.

Contributions. Articles submitted for publication should be addressed to Professor Peter Winch, Department of Philosophy, King's College, Strand, London, W.C.2. Contributors are asked to note the following requirements:

Articles should **normally** not exceed 3,000 words in length. **Occasionally,** however, longer contributions can be accepted;

They must be typewritten in double spacing on one side of the paper only;

Footnotes should be kept to a minimum and wherever possible avoided altogether;

Single quotation marks should normally be used, except for purposes of internal quotation and "scare" quotes.

Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

Contributors in the United Kingdom should enclose a stamped addressed envelope of suitable size; if immediate acknowledgment is required a stamped postcard should also be enclosed.

Overseas contributors who wish to have their MSS. returned should send an envelope **and international reply coupons** of the requisite value, whether for air or surface mail.

Proofs are normally read by the Editor. A contributor who wishes to read his own proofs should notify the Editor at the same time as the MS. is submitted.

Offprints. Contributors will be supplied free of charge with a limited number of offprints of their articles. Further copies may be ordered at a charge from the publishers. Orders for these, if required, should be placed **direct with the publishers** when an article has been accepted for publication.

The copyright of articles printed in Analysis remains the property of the author, but contributors are strongly advised, in their own interest, to consult the Editor before consenting to the reprinting of their articles.

PRINTED IN GREAT BRITAIN BY A. T. BROOME AND SON, 18 ST. CLEMENT'S, OXFORD

ol. 31 No. 4 (New Series No. 141) March 1971

ANALYSIS

Edited by
Peter Winch

CALC. UNIVERSIT

CONTENTS

Begging the question, 1971 — Richard Robinson

Empirical, analytic, evaluative — C. H. Whiteley

The *same* psychological state — Thomas C. Mayberry

Pretending and intending — Paul Helm

Can disembodied persons be spatially located? — Brian Smart

Concerning 'motive' and 'intention' — Burleigh T. Wilkins

What knowledge is not — W. R. Abbott

BASIL BLACKWELL · BROAD STREET · OXFORD

Twentyfive pence

ANALYSIS 31.4 MARCH 1971

BEGGING THE QUESTION, 1971

By Richard Robinson

TED. That begs the question.

JON. So what?

TED. What do you mean—'so what?'

JON. What does it matter if it begs the question?

TED. What does it *matter*? *It begs the question.* You shouldn't beg the question.

JON. Why shouldn't you beg the question?

TED. ?

JON. You don't know. Do you even know what begging the question *is*?

TED. Begging the question is assuming what you are to prove. For instance, if you are to prove that God is benevolent, you say 'God has all the virtues, therefore He is benevolent'. The premiss that God has all the virtues assumes that He is benevolent.

JON. It doesn't exactly *assume* it. It *entails* it.

TED. All right, then. It entails it. So it begs the question.

JON. So what?

TED. There you go again. 'So what?' indeed! *It begs the question.*

JON. You seem to imply that it is bad for a premiss to entail its conclusion. Is that your view? Do you think that a premiss ought *not* to entail its conclusion? And, when you condemn an argument by calling it a begging of the question, are you condemning it because the conclusion follows necessarily from the premiss?

TED. No-o; but—

JON. It is absurd, is it not? to condemn an argument because its premiss entails its conclusion. Rather we condemn an argument because its premiss does *not* entail its conclusion.

TED. Ye-es; but—

JON. I must say I do suspect you of this absurdity. I do suspect you of rejecting my argument just *because* my premiss necessitates my conclusion. At any rate, that is, in my opinion, what most people are doing when they condemn an argument as begging the question.

TED. Now I know what I mean. I mean that this single premiss by itself entails the conclusion, without the addition of any other premiss. The conclusion of an argument ought to be entailed by all the premisses taken together; but it ought not to be entailed by one of them alone, for that makes the others otiose. When it is entailed by one of them alone, I call it begging the question.

JON. So begging the question is the fault of including unnecessary premisses? Not a very serious fault, surely. It cannot invalidate the argument.

TED. No, that can't be it either. As you say, the addition of unnecessary premisses cannot remove the validity of a valid argument. I think begging the question is essentially the fault of arguing from a single premiss which by itself entails the conclusion, or even is equivalent to the conclusion. That shouldn't be done.

JON. You appear to be condemning what in my youth was called immediate inference. It was said that from 'All men are mortal' you could infer immediately that 'No immortals are men', but you needed a second premiss to mediate the conclusion that 'Socrates is mortal'. Another example of immediate inference was from 'No cats are dogs' to 'No dogs are cats'. Are you condemning immediate inference?

TED. Well, it's pretty useless, isn't it?

JON. I am sure there are some people who don't at once see that it follows. But anyhow, it's valid, isn't it?

TED. O yes, it's valid.

JON. So you are condemning a valid argument when you accuse an argument of begging the question?

TED. I suppose I am.

JON. You *know* you are. And it ought to make you hesitate to call an argument a begging of the question. For there just is nothing wrong about offering a premiss which by itself entails the conclusion.

TED. I still feel there is something wrong about begging the question, if I could only analyse it.

JON. There are only two proper ways of condemning an argument. One is to say that the conclusion does not follow from the premisses. The other is to say that you do not accept the premisses as true. Your begging the question appears to be neither of these. So it is not a proper accusation.

TED. But begging the question has always been thought to be a real fallacy. It's in Aristotle, isn't it?

JON. Yes, it first appears in Aristotle; but there it is breaking a rule of a game.

TED. *A game*?

JON. Yes, a game. The Academic game of elenchus, for two persons, the answerer and the questioner. The answerer asserts a thesis; and the questioner tries to refute him out of his own mouth, by asking him questions and using his answers to disprove his thesis. For example, if the answerer asserted the thesis that God is not benevolent, the questioner might ask him whether God is virtuous or vicious. If the answerer replied that God is virtuous, the questioner might ask him whether benevolence is a virtue or a vice. If the answerer said it was a virtue, the questioner might say: 'Come now, let us add our admissions together. We have said that benevolence is a virtue and that God is virtuous. It follows that God is benevolent, contrary to your thesis.'

TED. I don't see where begging the question comes into that.

JON. As in all games, there had to be certain rules, to make the game interesting or profitable. There were rules as to what the answerer might say, and rules as to what the questioner might say. One of the questioner's rules was that he might not directly ask for his conclusion. Thus, if he wanted to conclude that God is benevolent, he might not directly put the question 'Is God benevolent?' Nor might he put this question in a disguised form, such as 'Does the Supreme Being wish well to all creatures?' That is, he might not beg for the conclusion.

TED. But this is *beg* the *question*, not beg *for* the *conclusion*.

JON. 'Beg' means 'beg for' in this phrase, as in 'begging his bread'. I don't know why the English call it 'the question'. They are the only nation who do so, I believe. Aristotle called it 'the beginning', *τὸ ἐνάρχη*, and so did whoever translated it into Latin as *petitio principii*. And most modern nations merely repeat the Latin phrase.

TED. But the conclusion isn't the beginning either.

JON. It is in the best arguments. The best arguers, for instance Euclid, always tell you *at the beginning* what they are going *to conclude*. In the game of elenchus you always knew right at the beginning what the questioner wanted to conclude. By the purpose of the game he wanted to conclude the contradictory of the thesis which the answerer laid down. Since the answerer's thesis was the first move in the game, the conclusion desired by the questioner, namely the contradictory of the answerer's thesis, was known as soon as the answerer had spoken. Hence, if a questioner directly asked for his conclusion as such, it was natural to describe this as 'asking for the thing in the beginning'. It is a pity we don't say 'asking for' nowadays, instead of keeping the old-fashioned word 'begging'.

But you and I are not playing this game. We are trying to get at the truth, to know something. Therefore when I offer you an argument, you shouldn't tell me I am breaking a rule of the Academic game. You should consider whether my premiss is true, and whether my conclusion follows validly from my premiss. If you think my premiss is false, tell me so. And if you think my conclusion doesn't follow from my premiss, tell me that. Don't tell me that my premiss is true and my conclusion follows, but all the same my argument is bad because it 'begs the question'. That's absurd.

The prohibition of begging the question is not a law of logic, nor a maxim of good scientific method. It is merely a rule of an old-fashioned competitive game. To appeal to it when engaged in the scientific search for truth is as irrelevant as to obey the Queensberry rules when attacked by a murderer.

(*Two days later*)

TED. I've looked up Aristotle on begging the question; and I agree that in his *Topics* he regards it as a rule of a game. But he also says (*Topics* 162b 31–33) that that is not the *real* nature of begging the question; and he gives the real account of it in his *Prior Analytics* B 16, according to which it *is* a rule of scientific method.

JON. Yes, the mistake of trying to turn a rule of a game into a maxim of science was, I admit, begun by Aristotle himself. And I failed to mention this when we were talking the other day. But it is a mistake for all that; and Aristotle's *Topics* account of begging the question is better than his *Analytics* account of it, contrary to his own opinion.

TED. Why is it a mistake?

JON. Well, *something* must be wrong. That is clear from the fact that the two accounts are incompatible—unless indeed he meant different things by the phrase in the two works. But he didn't; and his *Analytics* account is a failure because it uses a concept which has no application in most of science.

TED. What concept is that?

JON. The concept of the self-evident, *τὸ διαυτου γνωστον*, that which is knowable through itself. *Prior Analytics* B 16 defines begging the question as 'trying to show through itself something that is not knowable through itself'.

TED. I don't think self-evidence is anything particular to do with begging the question.

JON. I don't think self-evidence is anything to do with science at all, except perhaps in mathematics and logic. In natural science nothing is self-evident; and no scientist tries to demonstrate any law of nature through itself. But Aristotle assumed or believed that all science when perfected would be like mathematics; and he plausibly believed that in mathematics there are self-evident propositions.

TED. So you think Aristotle's *Analytics* account is a failure, and his *Topics* account makes it merely a rule of a game which nobody plays any more?

JON. I am afraid I do.

TED. But the phrase is common in English reflective writing.

JON. I know it is. And I formerly used it myself. But now I think it is nearly always a muddle, or improper, or both.

Oriel College, Oxford

EMPIRICAL, ANALYTIC, EVALUATIVE

By C. H. WHITELEY

MANY philosophers have maintained that empirical statements (statements of fact) are to be distinguished on the one hand from analytic statements, which are true or false solely in virtue of the meanings of the words in which they are expressed, and on the other hand from evaluative statements, which express pro- or con-attitudes of some sort and are not properly speaking either true or false.

If this doctrine is taken to mean that all the significant assertions of ordinary language can be correctly placed in one or other of these categories, it is false. As regards the descriptive-evaluative distinction, the difficulty is that there are expressions of ordinary language which not only combine description with evaluation but do so in such a way that one cannot mark off one aspect of their meaning as primary and essential and another as derivative and inessential. If, at the bridge table, I say 'You have revoked', I at once describe your behaviour and condemn it as improper. Similarly, words like 'obscene', 'swindle', 'rude', 'bully', 'swank', 'neat', 'heroic', give in one breath both an evaluative assessment and its factual grounds. With sentences containing words of these kinds the inference from 'is' to 'ought' sometimes seems to be valid. 'Smith revoked' is implied by 'Smith, while playing bridge, played a diamond on a spade lead, having another spade in his hand', and also implies 'Smith is at fault'. 'Jones swindled me' is implied by 'Jones gave me a dud cheque in exchange for my car' and implies 'Jones did me wrong'. Of course, to accept these inferences is to accept the wrongness of failing to follow suit at bridge, and of giving dud cheques in exchange for cars. The point is that this acceptance is built into the common meaning of the words 'revoke' and 'swindle', and if the acceptance was not general there would be no such words. In a game in which following suit is optional there is no place for a word like 'revoke'. In a community which did not engage in buying and selling, or in one which was indifferent to commercial honesty, there would be no word equivalent to 'swindle'.

As regards the empirical-analytic distinction, the difficulty is of the opposite kind; it is not with sentences falling into both categories, but with sentences falling into neither. I take it that the meaning of a sentence is, for me, empirical if and only if I am prepared to accept some conceivable phenomenon as constituting a refutation of that statement; it is, for me, analytic if and only if I am not prepared to accept any conceivable phenomenon as a refutation of it. Thus, to interpret a statement as empirical or as analytic is to know what I should think it correct to say under certain hypothetical circumstances; and if such circumstances have never arisen I may not know what I should say if they did arise, and so may not interpret the statement either as analytic or as empirical, though I have a clear and adequate understanding of its meaning. I may know the meaning of 'All cats are furry animals', or 'A straight line is the shortest distance between two points', and yet not be able to say whether I should be willing to deny the former statement if an animal should turn up which was furless but otherwise catlike, or the latter statement if I came upon a line between A and B which looked straight to the eye, but measured a bit more on a tape-measure than another line between A and B which looked slightly curved. If I do make up my mind on these matters, I am not discovering a meaning which the expressions already had in my understanding, or anywhere else; I am determining how I shall use them in future. Before making up my mind, though I understood these sentences well enough, I understood them neither as analytic nor as empirical. Another man who understood them equally well might determine them as analytic or empirical differently from myself. Such equivocal examples are not scarce anomalies; in ordinary language they are very likely commoner than straightforward examples of analytic statements. In practical matters it is not often important to determine whether a statement is analytic or empirical so long as it can be relied on to be true.

This being so, the common philosopher's question 'Is this statement analytic or empirical?', and the related question 'Does the statement p entail the statement q?' are usually ill-advised in relation to non-technical discourse. If the question is of any interest, the answer is very likely to be indeterminate. It is not profitable for philosophers to inquire whether sentences like 'It is impossible for me to feel your pain' or 'Space has just three dimensions' or 'A cause must precede its effect' or 'We cannot see without eyes' convey empirical or analytic assertions. A determinate distinction does not exist in ordinary speech, and its introduction calls for a decision rather than for a report. What is worth doing in such cases is to describe the possible circumstances in which a decision would be called for.

So the traditional trichotomy will not do as a report about actual usage. If we are to take it seriously, we must take it rather as a recommendation about how language should be used. Now the indeterminacy between the analytic and the empirical arises because terms are not defined with sufficient precision. Once we have defined 'cat' and 'straight' precisely enough, the matter is settled. The empirical-analytic distinction would clearly hold throughout any language in which all terms were precisely defined in relation to one another; and to recommend that it should hold throughout a language is to recommend that all terms of that language should be so defined. The advantages of this are obvious. It is clear that in any theoretical enterprise we need to be able to distinguish between those general statements which are true as a consequence of our definitions and those which rest on empirical evidence and are therefore open to empirical refutation. And if in course of time it should become convenient to modify the definitions of our terms, we should know what we are doing and why we are doing it. Otherwise we may find ourselves involved with fallacious inferences in which one form of words is used sometimes in an empirical and sometimes in an analytic sense, or in disputes at cross purposes where one disputant interprets as sense what another interprets as nonsense. It is, however, worth noting that the vagueness of the vernacular has its own advantages. To make a meaning precise is to commit oneself in advance to certain linguistic behaviour which may turn out inconvenient when unforeseen anomalies or borderline cases arise. A rigid definition of 'cat' or 'democracy' or even 'straight' may hamper us by imposing inconvenient classifications on future material.

Why should we draw a sharp distinction between the descriptive and the evaluative? This would involve expelling from our vocabulary words like 'revoke' and 'offside' and words like 'swindle', 'obscene' and 'neat'. Such a wholesale prohibition seems unreasonable. It is not only in tendentious propaganda that such words come in useful. The honest citizen also from time to time finds it convenient to say in one word both that he approves or disapproves and why he does so. And we should have a harder task in introducing learners to our codes of behaviour if we could use only 'good' and 'naughty', 'right' and 'wrong' along with value-free descriptions, instead of having recourse to such question-begging words as 'polite', 'careful', 'selfish', 'tidy', 'loyal' and their opposites. The distinction is however needed in two kinds of context.

In the first place, if regulations are being made for people to obey, the things which they must do or forbear need to be precisely specified, in order that they shall know how to obey the regulation, and in order that breaches of it may be plainly recognised. This cannot be satisfactorily done in the sort of language which straddles the descriptive-evaluative gap. For the moral consensus on which such linguistic usages are based extends only to broad principles, not to details. We are agreed that what Dr. Crippen did to his wife was murder, because we are agreed that it was inexcusable. But whether shooting a burglar, poisoning an incurably sick relative, or killing one's child in a fit of insanity counts for us as murder depends on whether we think it excusable; and on this matter there are differences of opinion. Similarly, but more grossly, we are at odds or in doubt as to what constitutes obscenity. So a law which makes murder, or the utterance of obscenity, an offence leaves citizens in doubt as to what they may or may not do. To clear up the uncertainty involves defining 'murder' or 'obscenity' in neutral descriptive language. And then, as the connection of the words with some definite empirical criteria is tightened, their connection with moral evaluation is loosened, and it becomes possible to wonder whether some sorts of murder are really wrong. Defining our terms has the same effect with this distinction as with the other; it forces the difference between description and evaluation into the open. Clear regulations, then, need to be expressed in merely descriptive language. But clarity is not the only thing required of regulations, and this principle is never pushed to the limit. Every legal system contains expressions like 'due care', 'reasonable provision', 'gross negligence', 'grievous bodily harm', which contain an evaluative element. The reason for this sort of legislative vagueness is similar to that given in defence of vagueness in definition above. Circumstances and standards are variable, and it is wise to let magistrates take account of these variations in applying the law.

There is another context in which the distinction is needed, namely in the criticism and assessment of codes and standards. Rebels and reformers need it. The man who admits that a given work of art is a frank representation of sexual intercourse and denies that it is therefore vicious or reprehensible cannot operate with the word 'obscene' and cannot give a straight answer to the question 'Is this work of art obscene?' The trouble with words like this is that they hinder the expression of heterodox valuations—in a way, that is just what they are for. So, while they are in place when standards are being applied, they are out of place when standards are being criticised. So they are out of place in moral philosophy, in so far as moral philosophy is concerned with the criticism of standards, the justification of attitudes. The legislator needs the distinction between fact and right, between the question 'What did this man do?' and the question 'Is what he did wrong?' (This seems a likely source of the use of 'fact' by that erstwhile law student, David Hume.) The moral philosopher similarly needs the distinction between the question 'Is this right or wrong?' and the question 'Why is it right or wrong?' He needs to distinguish the expression of an attitude from a statement of the grounds on which this attitude is justified, and cannot work properly with a vocabulary in which they are confounded. Moral philosophy at least should be conducted in a terminology in which the deduction of 'ought' from 'is' is always invalid. It is worth remembering that it was in this context that Hume objected to it.

University of Birmingham

THE *SAME* PSYCHOLOGICAL STATE

By THOMAS C. MAYBERRY

SOME philosophers have made extensive use of a "strict" concept of identity in discussing psychological states. They draw a contrast between numerical identity and identity of type. For example, A. J. Ayer argues that:

> If one person is afraid and another shares his feeling . . . there are two feelings of fear and not one. To say that the feeling is shared is to say that the two feelings are qualitatively similar and that they have the same ostensible object; it is not to say that they are numerically identical . . . we are to say that there are two feelings and not one, just because there are two persons.[1]

The conclusion he draws from this is that 'it becomes a necessary truth that one person cannot have, and therefore cannot strictly know the experiences of another'. What I want to discuss is whether a paradigm of "strict" identity can reasonably be applied to psychological states. This question can be succinctly re-phrased for our purposes in this way: 'How do we go about counting psychological states?' And this question leads directly to: 'What are the criteria for a state being the same or a different state?' It seems that the answer to the question whether there is one state or two may be that it depends upon our purposes and the criteria which are relevant to those purposes. If, for example, we were counting birds, we could count each bird as one or each type of bird as one. But this answer is a little too easy, because it assumes either that we are quite free to adopt any criteria we choose in counting states or that the conventions for counting are the same for counting states as for counting birds. Assuming that we are interested in the criteria by which states are usually counted rather than in proposing new ones, we must look to see how such things are counted.

We do not count everything as if it could be fitted to a mathematical paradigm. Colours, shapes, heights, weights, and physical states, for example, are not counted in this way. If two mountain peaks, for example, are 14,000 feet high, volcano shaped, snow covered, and volcanically dead, we could say only as a joke that each had its own height, shape, colour, and state of volcanic death. The fact that there are two peaks does not tempt us to say that there are two heights or shapes. If they were different in these respects we could say this; otherwise not. When we count heights, colours, and so on, in the way we count objects we get paradoxical results. To such a question as 'Are these birds the same colour?' which may be taken to mean either 'Are they properly called by the same colour name?' or 'Are the colours indistinguishable?' we cannot appropriately answer that they are numerically different but qualitatively similar. 'Each bird has its own colour' assimilates colours to the wrong paradigm and so has a paradoxical ring. The same sort of paradox results when at least a number of states are treated in this way. For example, if two persons were sound asleep it would be paradoxical to say that there were two different but similar states in existence. If two birds were dead there would be two dead birds, but not two different but similar states of death. If there were two birds and each was singing its heart out, then there would be two birdsongs and two birds singing, but only one state describable as 'singing its heart out', and the birds would share that state rather than each owning a separate state.

---

[1] 'One's Knowledge of Other Minds', *Philosophical Essays* (London, 1954), pp. 194f.

To be in a state is not to own it. A bird, for example, owns its song but does not own its state, which may be that of singing. It owns its song in the sense that the song is uniquely identifiable as its song. 'Each bird has its own song' is a way of stating this fact. But 'Each bird feels its own urge to migrate', assuming that there is such a state, is confusing. Feelings are not uniquely identifiable in the way songs are and are not owned. Neither the individual who has a feeling nor an observer identifies a feeling as the feeling of a particular individual. 'Jim's feeling of fear' does not name a state. It simply describes Jim's psychological state. There are no individual psychological states any more than there are individual states of death or states of sleep. There is, in the latter case, sound sleep, restless sleep and so on, but none of these uniquely identifies or names the state of an individual. Similarly, in no case can we say that Jim's fear and Richard's fear are two different but similar individual states of fear. 'Jim's fear' no more identifies a unique fear than 'Jim's height' identifies a unique height, and Jim no more owns his fear than he owns his height.

The temptation to think of states as countable in the way in which individuals are counted may be partly due to a tendency to assimilate states to occurrences or happenings. We can and do say that these are exactly alike but numerically different. But states occur only in the sense that they come into being, persist for a period, and then end. During the period in which an individual is in a given state the state does not happen and is not happening. To be awake (or asleep), for example, is to be in a certain state but being awake is not a happening. Being awake is not something that happens to a person who is awake during the period he is awake. Similarly, being in a state of fear is not something that happens to him during the period that he is afraid. Contrast with this a rainstorm in which rain falls for a given period. Raindrops fall during any part of this period and the fall of each raindrop could be considered a separate occurrence. Because of this contrast between states and occurrences, we would never say of a rainstorm on Tuesday and a similar one on Thursday that the same rainstorm recurred, though we might say this of a toothache which was indistinguishable from a previous one. An occurrence such as a rainstorm can be counted in the way we count objects because it makes sense to name them for certain purposes, and bad storms do receive names. The same could be done for actions. For example, we might refer to a particularly memorable piece of bad behaviour on Richard's part as 'Richard's fit'. Even non-memorable occurrences can be referred to in this way. For example, 'the drizzle in Seattle on April 20, 1960' identifies a unique event even though it may be quite similar to many other drizzles. But 'Richard's fear' does not identify a unique occurrence, or happening, because fear is not an occurrence. What might happen is that the fear might grow until it was unbearable terror in which case we might identify the episode in which it existed as 'Richard's fit'.

Since a psychological state does not consist of separate occurrences which can be isolated, identified, and described, it can be said to recur after a period of time. Also, if two different persons are in a psychological state with the same description they can be said to be in the same state, to have the same feeling, and so on. A psychological state or a feeling is in this respect like the weather. Two widely separated cities can be said to have the same weather but not the same rainstorm. Two persons can share the same depressed state, have the same rheumatic pains, be the same height, and the like. Two cities can have the same weather, topographical features, size, and the like. They cannot have the same rainstorm or plane crash even if the descriptions of these are identical. These are different occurrences, have different locations, and so on. But the weather is not an occurrence and is not located anywhere, and the same is true for a psychological state. Though a person sleeping has a location, his sleep has no location. This is true for depression, fear, and many other states; though pains, tickles, tingles, and so on, can be located, which increases the temptation to treat them as occurrences. But being in pain is nevertheless not a happening. What happens is that one comes to be in pain or it grows more or less or it stops. If being in pain were a happening, it would not only be possible to locate it but to identify various stages or phases in the happening, and this cannot be done.

States can vary in degree from very mild to extremely severe and the transition from one degree to another can be regarded as an occurrence. And these can be counted in the usual way in which occurrences are counted. For example, one might have two severe attacks and one short period of comparative relief during an extended period of being in pain. Also states such as being in pain, having a tickle, and the like are counted as different when their locations differ. For example, if there is a pain in my back and another in my knee, I have two pains. But if there is a pain in Jim's back and a pain in the same position in Richard's back, this does not count as a difference in location, and whether or not they are considered to be different will depend upon their descriptions. Pains which differ in their descriptions are counted as different. Those whose descriptions coincide are counted as the same. If each of two persons can feel a dime in his pocket, then there are different dimes which each feels. But if each of two persons feels a twinge in his back, there are not necessarily two different twinges which each feels. Whether or not they feel the same twinge depends upon the description of the twinge. This is a respect in which twinges differ from dimes and in which 'feels' in the one case differs from 'feels' in the other. Twinges are not objects and 'feels a twinge' means the same as 'has a twinge'. The inclination to count twinges as if they were dimes or happenings such as rainstorms and 'feels' in the sense of 'has' as if it were 'feels' in the perceptual sense is responsible for a great deal of the confusion concerning the way states can be counted.

The attempt to identify states and count them by reference to their "owners" is another source of confusion. Strawson, for example, has attempted to count states in this way and writes as follows: 'States or experiences, one might say, *owe* their identity as particulars to the identity of the persons whose states or experiences, they are'. He concludes from this that it becomes 'logically impossible that a particular state or experience in fact possessed by someone should have been possessed by anyone else'.[1] But this is like saying that it is logically impossible that the particular height, weight, shape, appearance, and the like, "possessed" by one person could be "possessed" by another. Or that the particular weather possessed by one city could be possessed by another. But cities do not own their weather with a 'logically non-transferable kind of ownership' any more than I, for example, own my putt or my drive when I play golf. 'My' putt is just the putt that I will sink or fail to sink or have sunk, *etc.*, from the position where my golf ball lies on the green. One can own golf balls but not putts. Hence a 'logically non-transferable kind of ownership' of putts, weather, and states is no ownership at all, and the belief that it is is based on a misinterpretation of the possessive case in which it indicates not "ownership" of states or actions, but actions performed by me or states in which I happen to be.

Persons do happen to come to be in certain states so what can be identified as a 'particular' where this is necessary is a person in such a state. The state cannot be identified as the state in or of a particular person, because states are not named or described in this way. Even if there were a state, as there is a disease, called Parkinson's state it would be a state having a certain description in which anyone might come to be. To uniquely identify a state is simply to describe the kind of state it is. States are not classes made up of particular individuals, as the class of orioles is a class so made up. Therefore, the problem of identifying particular individuals within the class never comes up. 'Richard's pique' no more identifies a particular individual within the class of piques than 'Richard's weight' identifies a particular individual within the class of weights. There is no such class in either case. What each of these identifies is the individual who is piqued and a certain weight. And two individuals can be in the same state of pique just as they can be the same weight. To be in the same state of pique as Richard is simply to be piqued in a way having the same description in the relevant respects as the description of Richard's pique. Treating a state of pique as if it were an individual having certain properties, one of these being correctly described by the phrase 'belonging to Richard at *t*', does result in the 'logically non-transferable kind of ownership' which Strawson attributes to such states. For by Leibniz's Principle of the Identity of Indiscernibles *x* and *y* are identical if and only if every property of *x* is a property of *y*, and no state belonging to any other person could have the property of belonging to Richard. But, as I believe I have shown, states are not individuals and 'belonging to Richard' is not a property of Richard's states or his height, weight, shape, or the like. Hence, Leibniz's principle is not applicable.

---

[1] 'Persons', *Minnesota Studies in the Philosophy of Science*, ed. H. Feigl *et al.* (Minneapolis, 1958), Vol. II, p. 336.

To remind ourselves how serious and important the question whether something or someone is the same or different can be, consider the following imaginary case. A tribe exists in which adult males only retain their identity so long as they are wearing a tattoo which they received during their initiation and naming ceremony. Identity assassinations sometimes occur in which the tattoo is removed. Individuals so 'de-identified' lose their names, their identity, their tribal rights, and so on. They can acquire a new identity by a petitioning procedure but are never treated as the same person as before in legal or social transactions, are not considered to have a family but must acquire a new one, and so on.

Contrast this case with what William James says about states: '*No state once gone can recur and be identical with what it was before*' (his italics) and 'However we might in ordinary conversation speak of getting the same sensation again, we never in strict theoretic accuracy . . . do so'.[1] In the tribal identity example there is a clear contrast between being the same and being different; criteria are spelled out by means of which a decision can be reached as to which is the case. But if we follow James' lead with respect to psychological states any basis for a meaningful contrast is ruled out, with the result that 'same' and 'different' no longer have a job to perform in this context. No criteria exist by which to determine whether two states are the same or different.

---

[1] *Principles of Psychology* (New York, 1896), Vol. I, p. 230.

James' attack on the ordinary criteria by which decisions are reached concerning the sameness and difference of states may have many roots. Perhaps a vaguely felt criterial looseness is the source of a great deal of this dissatisfaction. But why does it seem so important to use 'the same' so precisely and strictly? Here I think the answer is that an inapplicable model of strictness suggests itself and that 'strictness' on this interpretation just is taking that model as the paradigm for states. So the strictness is really a pseudo-strictness derived from applying the wrong criteria. And the answer to the question is that it just seems very important to apply that model of strictness to states, because it is not realized that there is more than one model for what counts as the same. When it is thought that 'the same' means what it means no matter what the context or subject matter, *i.e.*, that it means what that model tells us, then no other model can suggest itself as even plausible.

But when it is realized that the criteria for 'the same' depend upon the context, and that the ways in which things are determined to be the same in those contexts tell us whether two things are the same, a move to adopt a one model meaning for 'the same' for all contexts should be seen to be not a move toward "strictness" but a move toward unnecessary and confusing criterial revision.

University of Toledo

PRETENDING AND INTENDING

By PAUL HELM

THIS paper defends the idea, rejected both by Austin and Miss Anscombe,[1] that pretending necessarily involves the idea of limits, or, as Austin puts it, that pretending entails not really being, and really being entails not pretending. I distinguish three cases of pretending, and argue that the two implications hold for all three, and that Austin's view is only plausible because he fails to make certain distinctions, and wrongly describes certain examples that he cites. This argument does not amount to a complete defence of the limit view of pretending, as there is a class of examples in which the implications clearly do not hold, those which involve certain false beliefs that the pretender has. For example, a man may pretend to be the oldest man in Liverpool, or to be poisoned (Miss Anscombe's example), and be unaware that he is in fact the oldest man in Liverpool, or poisoned. But this class of cases only holds for pretend$_1$ (see below).

---

[1] In their symposium 'Pretending', *PASS* 1958, Austin's contribution reprinted in his *Philosophical Papers*, Oxford 1961.

I

The three cases of pretending are:

1. A pretends$_1$ when he acts in order to deceive$_R$ B, where deceiving$_R$ involves getting B to believe p, where p is some proposition that A knows to be false, but which is not known by B to be false, or (more weakly) is believed by A to be false and not known by B to be true. A will get B to believe p by (typically) behaving in certain ways, including saying certain things, but his intention to get B to believe p is opaque to B, at least in cases where the pretence succeeds. An example would be A pretending to have been at home by saying/doing certain things in the presence of/to B. B's believing that A was at home as a result of the behaviour of A is a sufficient condition of B's being deceived by A (though B's belief that A was at home is by itself not a sufficient condition of deception, for that may just be a case of B believing something falsely). A acting to deceive B does not necessarily exhaust A's intention in the matter (though it may, where the pretence is serious, but has no further end. A could act to deceive with the further intention of hiding the whereabouts of another person, and his motive could, for example, be patriotism or cowardice.)

2. A pretends$_{11}$ when he acts in order to deceive$_C$ B, where deceiving$_C$ involves getting B to recognise that A has this intention. A's intention to get B to believe what is false is transparent to B (in successful cases). A will do this not only by behaving in certain ways and/or saying certain things, but also by occupying a particular conventional setting. Examples would be a stunt-man, or certain kinds of film, or Austin's case of a conjuror pretending to saw a woman in half. In the normal case it is not that the conjuror gets the audience to believe that the woman is sawn in half; otherwise they would react rather differently, perhaps rush on to the stage, have the man arrested *etc.* Rather, the conjuror gets the audience to believe both that the woman appears to be sawn in half, and is not sawn in half, at least where the act is successful. If the conventional setting is not grasped (*i.e.* in this case that the man is on a stage, giving twice-nightly performances, *etc*), then pretend$_{11}$ collapses into pretend$_1$ at least as far as the beliefs of the audience are concerned.

The deception here is conventional (deception$_C$) because it allows the recognition of the performer's intentions. But the distinction between conventional and non-conventional is not a hard and fast one, and cases can be imagined in which because of this it is not clear whether the action is a case of pretence$_1$ or pretence$_{11}$.

3. A pretends$_{\text{III}}$ when he acts in order to appear to B to be what he necessarily is not, or to do what he is necessarily not doing, with no intention to deceive as to who he is, or what he is doing. A human being is necessarily not a kangaroo, and except in pathological cases knows that he is not. So that if he pretends to be a kangaroo this is a case of appearing to be what he necessarily is not, a case of pretence$_{\text{III}}$. The fact that there is no intention to deceive does not entail that the action is a case of what Austin calls 'let's pretend' *i.e.* purely frivolous, for the action could be part of some larger non-frivolous deception. Though A has no intention to deceive as to who he is his purpose in pretending to be a kangaroo might be to distract B while C rifles B's pockets.

It might be argued that the notion of 'producing an illusion' depends on the notion of deceit$_{R}$. What this shows is that the cases of pretence given above are related, but not that they are not distinct. It is not intended that the three cases are exhaustive. A rough way of distinguishing between cases 2 and 3 would be to say that case 2 could have been a case of 1 (if the conventions were absent, or ignored, or not recognised), but that 3 could never be a case of 1.

The success of a pretence will be relative to the sense of 'pretend' concerned. There is, for example, the distinction between the success of a pretence as a performance (our senses 2 and 3) from the success of a pretence as a deception. When Austin talks of the ease or difficulty of pretending[1] he is talking not about pretending *tout court* but about the ease or difficulty of *succeeding in deceiving.*

Austin's failure to distinguish between different senses of pretend leads to other mistakes. He is plainly wrong in saying that we should never say of a small boy in an armchair making tugging and twisting movements with his arms, *etc.*, that he is pretending to be driving a racing car, on the grounds that there is 'no serious prospect of deception' in this case. Granted that there is no such prospect, this could be a case of pretend$_{\text{III}}$. And what he says of another example is similarly implausible. 'Trapped on a branch against the moon, we decide to pretend to be owls: you give a colourable hoot while I pull up my legs and hunch my shoulders to my ears' (p. 213). He says of this 'you stop short of pretending to be an owl, because you fail to attempt to disguise the fact that you are not one—mere imitation does not imply dissembling anything'. This is false on two counts. If what is described is a case of pretending$_{\text{III}}$ then the necessity of attempting to disguise the fact does not arise.

If it is a case of pretending$_{\text{I}}$ then there is no means of deciding *a priori* what will be sufficient to constitute disguise. If all that is necessary for disguise is auditory deception, then hooting will be enough. When Jacob deceived blind Isaac by pretending$_1$ to be Esau all the deception that proved to be necessary was tactual.

---

[1] 'It is easy to pretend to be sitting on a certain chair when it is half-concealed behind a desk, less easy if it is in full view' (p. 207—all page references to Austin are to *Philosophical Papers*).

Because of the distinct senses of pretending there can be no essence of pretending, as Austin claims (p. 210). To say 'public behaviour is meant to disguise some reality' trades on the ambiguity of 'disguise'— a person can disguise something without having the intention thereby to deceive —and on the ambiguity of 'is meant to' which can cover conventional as well as non-conventional cases. To say 'To be pretending, in the basic case, I must be trying to make others believe, or to give them the impression . . . that I am (really, only, *etc.*) *abc*, in order to disguise the fact that I am really *xyz*' (p. 215) is to roll together three distinguishable cases. To give B the impression that p is not the same as to get B to believe that p. Nor if A gives B the impression that p is he necessarily disguising the fact that not-p. Nor if A gives B the impression that p is he necessarily intending to disguise the fact that not-p.

What are the relations between the three senses of pretend and the two implications mentioned earlier, that pretending entails not really being, and really being entails not pretending? At first sight it might seem that they require modification in the following way:

1. A's pretending to ϕ entails A's belief that A is not ϕ-ing.

2. A's belief that A is ϕ-ing entails A's not pretending to ϕ. That is, it might seem that there must be a reference to the beliefs of the agent to cover the cases mentioned earlier in which a person may pretend to be C or to ϕ when he believes he is not C or not ϕ-ing, though in fact he is. (He believes he is not the oldest man in Liverpool, and pretends to be, though all the while he is.) But this further condition does not seem to be required in the case of pretend$_{11}$ and pretend$_{111}$. In the case of pretend$_{111}$ a necessary condition is that the agent knows that he is not the one he is pretending to be, and in some cases knows that he is necessarily not the one he is pretending to be, *e.g.* an animal or a tree or Napoleon. Hence since knowing p entails believing p, someone who pretends$_{111}$ to ϕ necessarily believes that he is not ϕ-ing. And in the case of pretend$_{11}$ the same seems to apply, for the agent in this case wishes his intention to give the appearance of ϕ-ing, *e.g.* sawing the woman in half, to be recognised. (I neglect the strange case of the conjuror who *actually* and intentionally saws his partner in half on the stage as an unnecessary complication.) This suggests that the action concerned must be intentional, and hence that the agent cannot be mistaken in his beliefs about whether he is doing this action or some other. But the further condition is required in the case of pretend$_1$.

Austin's argument to the effect that the implications do not hold in certain cases of pretend consists of two examples which I now want to discuss. The first is as follows:

> Two miscreants are surprised in the act and hastily agree, the wherewithal being handy, to pretend to be sawing a tree: in a trice the blade is humming to and fro a bare inch away from the bark. How good a pretence is that? And wouldn't they any longer be pretending to be sawing the tree if they allowed the teeth to bite in? Surely if they want the pretence to be convincing they should set about actually sawing the tree? (p. 207).

This is the argument for the claim that pretending to ϕ is in some cases compatible with ϕ-ing. But it does not show this. Rather it shows that ϕ-ing at t_2 can be a pretence as ϕ-ing from t_1-t_3, which I take to be the point behind Austin's distinction between pretending to saw the tree, and pretending to be sawing the tree. A person cannot pretend to saw a tree by sawing a tree though he can pretend to have been sawing a tree for the past hour by sawing a tree, or pretend to be a woodsman by sawing a tree for an hour. In the example the men are not pretending to saw the wood; they are sawing the wood and hoping thereby to deceive, which is something different. Austin has misdescribed a case of pretend_1; it is not that these are cases of pretend_1 for which the implications do not hold. Austin's other example is as follows:

> That chap over there, he's all right I suppose, he's cleaning the windows, eh?
> Ah, *him*, he's *pretending* to be cleaning the windows right enough, cleaning 'em a treat too: but I seen him taking note of the valuables through 'em all the time (p. 207).

Austin's comment on this is

> we must allow that he is indeed actually cleaning the windows, from start to finish and throughout the whole time he is pretending to be cleaning them. But it is still a pretence, because what he is *really* doing all the time is something different, namely noting the valuables: he is only cleaning the windows to disguise and promote this other activity (p. 210).

I do not accept this. It is not a case of pretending at all, but of doing one thing in order to promote another. It is not that the man is really noting the valuables, and not really window-cleaning. He is doing both really. Neither 'He is pretending to be cleaning the windows' (if this is meant to characterise an activity over a length of time) nor 'He is pretending to clean the windows' describes this case adequately.

II

It might be thought that Miss Anscombe's distinction between real pretences and mock performances is sufficient, and that the three-fold distinction made earlier is unnecessary. But her distinction will not do for a number of reasons.

1. If there are different sorts of pretence (as I have argued there are) it is not helpful to distinguish 'real' cases from others. Pretend$_{1}$, $_{11}$, and $_{111}$ are all cases of 'real' pretending.

2. Her distinction suggests that non-real pretences (by which she must mean cases of pretend$_{111}$ are necessarily performances. She says that mock performances are most naturally exemplified in 'current personal performances' in the presence of others, but (p. 279) gives no argument for this.

3. Her description of real pretence as 'trying to appear what you are not' covers cases of mock performance as well. 'Appear' is ambiguous; someone can appear to be x, and be not-x; and appear to be x, and be x. Further the 'mock performance' could refer to the intention behind the action *i.e.* if there is no intention to deceive then the pretence is a mock performance, or it could refer to what is done. Thus pretend$_{11}$ and pretend$_{111}$ could both be thought of as mock performances, though they have markedly different features.

4. The distinction suggests that only real pretences can involve the intention to deceive. But mock performances can, as we saw, though the intention to deceive is a further intentional consequence.

Miss Anscombe's argument against the 'limit' view of pretending is to show (rightly in my view) that, for example, being angry involves more than behaviour. It involves having characteristic aims, being essentially directed to certain objects. But why is this thought to count against the limit view of pretending? Given that anger involves more than behaviour a man can give a limitless, flawless piece of anger behaviour, and not be angry. But the limit comes when he stops short of fulfilling the characteristic aims of an angry man, for example the destruction of the object of his anger. If to be angry at x is intentional, then to pretend to be angry at x will either involve giving a flawed piece of anger-behaviour, or the placing of a limit on the fulfilling of the intention in some respect, or both. This does not show that pretending to be angry does not entail not really being angry; only that pretending to be angry is compatible with behaving angrily.

University of Liverpool

CAN DISEMBODIED PERSONS BE SPATIALLY LOCATED?

By BRIAN SMART

DOES it make sense to assign to a disembodied person spatial location? I shall examine various forms of an argument[1] which is designed to show that it does make sense and which provides a non-corporeal criterion for the location. I shall attempt to refute the argument and shall argue that *if* disembodied existence is to be defended then it would be essential that spatial location were not a feature of it: whether disembodied existence could be defended in its own right is beyond the brief of this paper.

Before turning to the argument it might be worth dismissing two quick ways of dealing with our problem:

(1) One way would be to supply a bodily surrogate, for instance a segment of the brain, a puff of smoke, a patch of light or shade; if these surrogates were allowed, spatial location would have to be assigned; however this line is unsatisfactory as there is clearly a sense in which the person is still 'embodied' or 'enmattered';

(2) the second way suggests that, by definition, 'disembodiment' precludes 'spatial location'; an appeal might be made to the Platonic and Cartesian concepts of the soul; yet what is required for this *a priori* thesis is an argument since there are rival *a priori* theses in the field; as we shall see, there is not just one neat conceptual step from 'X is disembodied' to 'X has no spatial location': various strands involving concepts of identity go to make up the thread.

The Argument

Consider how we infer the position of a television camera from the sequence of pictures it gives on the screen. While we could in principle check up on the correctness of our inferences (by actually looking at the camera itself) we hardly ever do. Similarly, we often infer our own bodily position by looking around our environment without actually seeing our bodies at all. In its simplest form the argument maintains that a disembodied person must locate himself on this kind of evidence alone. Obviously the person cannot check up on the correctness of his inferences by looking at his body, for there is none.

We can refine the argument in three ways.

---

[1] Terence Penelhum produces an argument along these lines in *Survival and Disembodied Existence* (Routledge and Kegan Paul, 1970) Chapter 2. He contends that if problems of identity could be overcome there is a non-arbitrary criterion for location (p. 25); later he correctly denies (e.g. pp. 70–1) that these problems can be overcome. My argument is that even if they could be overcome no sense could be attached to the claim that a disembodied person occupies space.

(1) There is no sense in talking about inference and evidence as if some hypothesis has been constructed: the viewpoint of the disembodied person is not on this view *evidence for*, but *the criterion of*, his location. The embodied situation is quite different, for there we do have a checkable hypothesis. The viewpoint in the pictures on the television screen might be the centre of the room and, in the normal course of things, the camera which is transmitting the pictures would be found in the centre of the same room. It is also logically possible that though the viewpoint of the pictures transmitted should be the centre of the room, the camera itself should be in the corner: there might—or there might not—be an explanation for this; but if we were satisfied that the pictures on the screen did depend on the correct functioning of the camera we should have no choice but to say that it was the camera that was transmitting the pictures: we might find that the pictures disappeared when the camera stopped working; that when the camera moves round 360° the pictures on the screen are those that would normally be transmitted by a camera (in the centre of the room) moving round 360°.

In the case of embodied persons it is a contingent truth that their visual point of view coincides with the location of their eyes. Yet it is logically possible not only that their viewpoints should be located on some other part of their bodies but that their viewpoints should also be located at some point away from their bodies. Take the following case: a man constantly sees objects from a point of view ten yards to the left of his body; when he wants to look at objects in front of his body he either sees them to the right of his visual field if he keeps his eyes and head facing frontwards, or, if he turns his head half right he sees them in the centre of his visual field; if he wants to look at his own body he turns his head 90° to the right and sees ten yards away the left side of his body and the back of his head; if he turns his body 90° right then his point of view will now be ten yards to the left of his body in its new position or, what amounts to the same thing, ten yards in front of his body in its original position; whether or not he sees anything at all depends upon whether he has his eyes open and so forth. I suggest that here there would be good grounds for saying that the man does use his eyes to see, only the point of view does not coincide with his eyes. (I agree with Sidney Shoemaker in chapter 5 of his *Self Knowledge and Self-Identity* (Cornell University Press 1963) that it is logically necessary that one's point of view should not be constantly shifting as this obliterates 'the distinction between egocentric statements and other kinds of statements, and with it the distinction between perceptual knowledge and clairvoyance' (pp. 181–2). But I fail to see how he can use this truth to establish that the viewpoint should be located somewhere on the surface of the percipient's body. My imaginary case has the merit of involving the use of one's eyes and head, a feature which Shoemaker regards as dispensable in cases where one's viewpoint is located on the base of one's skull or one's shoulderblades.)

(2) It would be odd to identify a disembodied person's location by the viewpoints contained in *all* of his perceptual experiences. If he is to be said to be *perceiving* at all, rather than just having a series of experiences, it is essential for him to be able to distinguish between veridical and non-veridical experiences. He would be able to construct a coherent space-time world from his experiences and class as illusory or hallucinatory any discordant elements. While these discordant elements would contain points of view he would not identify his location by them. Unless this feature of embodied experiences is imported into the disembodied situation there seems little point in assigning to the disembodied person spatial location—location in a public space.

(3) The most plausible case for the idea of disembodied persons holds that they were formerly embodied. The following possible experience would give great weight to the idea that such a person, though disembodied, still occupied space: an embodied man looks down and sees his body rapidly disappearing: finally his whole body is like a 'phantom limb', not being visible or tangible or audible, nor presenting any resistance to objects passing through it. During the whole process the man has continuously observed the room which his body was occupying until the body disappeared; he concludes that, though disembodied, he is to be located where his body formerly was.

Assessing the argument

The argument we have to assess is, it seems, in defence of the *a priori* decision to identify the disembodied person's location with his general point of view. In the absence of a body this might be felt to be theoretically the most economic anchoring point left. But is it? Take a man who, while embodied, always experienced seeing and hearing things from a point ten yards to the left of his body, but who also felt pains, tingles and tickles where they are normally felt—in his body; if he now sees his own body disappear he is likely to locate himself in the region of his pains, tingles and tickles, not in the region of his perceptual viewpoint.

How could we arbitrate between these two accounts of the disembodied man's location? They are not rival hypotheses since they cannot be tested by an independent check on bodily location as there is none: rather they supply *a priori* rival criteria for personal location, the first citing perceptual viewpoint, the second "bodily" sensations.

It may be suggested that what is underlying these rival theses is a tacit appeal to the bodily criterion: normally one's viewpoint coincides with the location of one's eyes and ears; where this is so we can, it may be suggested, locate the person by his viewpoint once his body has disappeared. On this suggestion it seems to follow that if a man has always had his viewpoint ten yards to his left then he is to be located ten yards to the right of his viewpoint once his body has disappeared. The reason why the man was located by his bodily sensations and not his viewpoint in the example above was simply because his bodily sensations had always been located in his body: if his viewpoint had been located on his body and his sensations had been located at some point from his body (as in cases of phantom limbs) then his viewpoint would have supplied his location.

The difficulty with these theses is that they do not take into account shifts of viewpoints and sensations. Take an embodied man whose viewpoint is located on his eyes and ears at T1 but finds at T2 that his viewpoint is one yard to his left; his viewpoint continues to shift until at T10 his viewpoint is ten yards to the left of his body; he still locates his sensations in his body from T1 to T10. Now suppose it were held that this was what could happen, in a sense, to a disembodied man? That is, that while *he* is still located, stationary, where his bodily sensations are, his viewpoint has shifted from where he is to a region ten yards to his left. What would be the difference between this description and one which claimed that while his bodily sensations continued at the original place, *he* moved ten yards to his left? If there were a body there would be a difference, but without a body there can be none.

Even if we widened the class of bodily sensations to include kinaesthetic sensations, body image and the so-called non-observational knowledge of one's body, we should have no way of settling the disembodied case. For suppose I were to discover that my voluntary movements of the phantom limb were to have some corresponding movements in my real, but severed, limb: then, while my limb does not occupy the space that my phantom limb does, there is a sense in which a part of me is located where my severed limb is: I am located where my body is, not where my body image or phantom limbs are. So even if my shift of viewpoint is accompanied by my body image voluntarily walking ten yards to the left this is perfectly compatible, in the embodied situation, with my remaining where I am.

Suppose the defender of disembodied location simply claimed that where there was a separation of viewpoint and sensations the question was undecidable, but that when they coincided it was decidable. He then faces the difficulty of the body's reappearance: if a qualitatively similar body appears where his viewpoint and sensations are located he will presumably identify this body with the one that disappeared: but suppose that another similar body appears at precisely the same spot where the old one disappeared: how could we decide between the possibility (a) that the new body at the original spot was identical with the old body, while the viewpoint and sensations had simply shifted; and

(b) that the new body located where his viewpoint and sensations are is identical with the old body? If we decide in favour of either (a) or (b) we are presumably deciding whether the man has moved or remained stationary. Yet unless his body can be ascribed continuous existence I do not see how we could in principle decide the issue.

Another possibility to be examined is that where the viewpoint and sensations diverge the person is to be located at their divergent locations. (The corporeal analogue of this is the person who still has voluntary power over his severed limbs.) But if both (a) a body like his old body reappeared at its original location and (b) pieces like pieces of his old body appeared at the divergent locations of his viewpoint and sensations, is the body in (a) or are the pieces in (b) to be identified with his old body?

The trouble with any discontinuity in bodily existence is that it immediately allows of the logical possibility of reduplication. But even if we could establish identity through discontinuity it is not clear how location either before or after disembodied existence is supposed to bear on location even inductively during disembodied existence.

All paradoxes about location for disembodied persons, though not all problems about identity, could be resolved if we refused to assign them spatial location. This is the course taken by Descartes and the non-mythologising Plato. But what conceptual reasons could there be to defend the thesis that disembodied persons have spatial location? There is, I think, a legitimate concern with the problems of identification and individuation. If disembodied persons have no location, then it is logically impossible for two exactly similar sets of experiences to coexist. This problem can be overcome if the owners of the experiences can be spatially distinguished: clearly this can be done if they have distinct bodies and even if the bodies are exactly similar. If disembodied owners of experiences could be spatially distinct then this problem can be overcome. If it cannot, then we are faced with the question of what kind of identity it is that we can ascribe to such persons when in the theoretically crucial cases identity cannot be distinguished from exact similarity.

The relevance of spatial distinctness to numerical identity may be brought out like this: no two objects can occupy exactly the same space: if *per impossibile* they do, then they are one and the same object. The question now arises: can two disembodied persons occupy exactly the same space? If they can then they are identical; and certainly if they were originally distinct and then converged upon the same place no criterion could possibly identify the two souls that then diverge from that space with either of the two original converging souls. If they cannot, then they are not really disembodied: for what could possibly prevent them from converging? If they actually met resistance as they attempted to converge they would surely be two physical substances of some kind.

It follows that if disembodied persons were ascribed spatial position this would be an idle ceremony: no standard conceptual work would be done by this ascription. As it is, no paradox-free account can be given of their location even if real conceptual work were done by locating them: any criterion for location is arbitrary and is always in competition with other—equally arbitrary—criteria.

Are there any other motives for attempting to spatialise disembodied persons? In addition to the feeling of continuity there is perhaps one, and a very pervasive one. Many languages tend to assimilate existential and locative uses of the verb *to be*: thus *to exist* tends to be interchangeable with *to be in such-and-such a place*. Charles Kahn[1] has provided strong evidence for this in ancient Greek: the non-mythologising Plato showed the pressures of his language and his powers to resist them when he assigned the Forms and souls to the *noetos topos*, the intelligible region outside of space. But neither Plato nor Descartes foresaw the now familiar difficulties in providing criteria for identifying and individuating non-spatial substances whose existence is somehow continuous with corporeal persons. What I have done is to suppose that there should be experiences of the kind considered above and have attempted to evaluate them in relation to the question of location: whether they supply adequate material for identifying persons is not a question I have gone into, though I have done enough to indicate my prejudices.

---

[1] 'The Greek verb *to be* and the concept of Being', *Foundations of Language* (1966), Vol. 2, pp. 245–265.

University of Keele

CONCERNING 'MOTIVE' AND 'INTENTION'

By BURLEIGH T. WILKINS

WHILE progress has been made in recent years in the analysis of 'motive' and 'intention' separately, there has been little progress in the systematic exploration of the resemblances and differences between 'motive' and 'intention'. Miss Anscombe's remark that an intention is 'a forward-looking motive',[1] although suggestive, isn't spelled out; and her effort to distinguish forward-looking motives from backward-looking motives and motives-in-general encounters the difficulty, apparent in her own examples, that too many motives are complex mixtures of two and sometimes of all three of the kinds of motives she has listed; moreover, her criticism of Ryle's dispositional analysis of 'motive' obscures an important sense of 'motive' and is, I think, mistaken as to what a dispositional analysis of 'motive' actually involves.[2] I hope to show that where one sense of 'motive' is concerned her suggestion that an intention is a forward-looking motive is in general correct and yet in one respect misleading; and it is this sense of 'motive', I believe, which is at once the most fruitful and the most troublesome when we want to compare and contrast 'motive' and 'intention'.

But are we certain that we want to distinguish between 'motive' and 'intention'? Often though not always the two terms seem to be used interchangeably, and the same answer often seems to do as well, or nearly as well, whether our question is phrased in terms of motives or intentions. Considerations such as these tempt the Pragmatist Within to wonder on those occasions when 'motive' and 'intention' do not interchange easily, or at all, whether anything of philosophical significance (as distinct from the nuances of an informal language) is involved. But, of course, there is the Realist Within warning us not to ride roughshod, as the Pragmatist is inclined to do, over differences in language, at least not until we know why there are differences, in this case why 'motive' and 'intention' aren't always interchangeable.

Miss Anscombe notes that 'popularly, "motive for an action" has a wider and more diverse application than "intention with which the action was done" ' (loc. cit.,) p. 19. This is the point at which I wish to begin considering some of the relationships between 'motive' and 'intention'. I believe that 'motive for an action' has a wider and more diverse application than 'intention with which the action was done' because, to put my thesis in its simplest form, while there are three senses of 'motive' that are relevant here, 'intention' is interchangeable with only one of these three senses and then not always.

---

[1] G. E. M. Anscombe, *Intention* (Oxford, 1958), p. 21.

[2] See my ' "He boasted from vanity",' ANALYSIS (April, 1963), ff. 110–112.

'Motive' can, I think, be used (1) to mark the presence of a reason for acting, (2) to mark the presence of a reason for acting *and* to indicate that this is in fact an agent's reason for acting, and (3) to mark the presence of a disposition in an agent for acting in a certain way under certain kinds of circumstances. (2) corresponds with what is sometimes called 'the occurrent sense of motive' and (3) corresponds, of course, with Gilbert Ryle's dispositional sense of 'motive'.[1] What I have in mind by (1) can perhaps be brought out by the following example: at the beginning of a murder investigation a police detective learns that Jones stood to benefit from the will of the deceased and subsequently says to Jones, 'You had a motive for murder'. At this stage in the inquiry, this need only be a way of saying to Jones that he had a reason to commit murder and it is not to make an accusation, formal or otherwise; ordinarily, an accusation must wait until the detective believes both that there was a reason for murder and that this was in fact the accused's reason for committing murder. (I say ordinarily not always because there are the so-called senseless murders "done without reason".) Both (2) and (3) seem to me legitimate (and not, as Ryle believed, mutually exclusive) senses of 'motive'. In the case of (2), to speak of the occurrent sense of 'motive' implies that we can usually date an agent's decision to act on the basis of a given reason. ('At twelve o'clock he learned he was mentioned in Smith's will, and that afternoon while walking in the garden he decided to murder Smith.') 'Motive' in sense (2) ordinarily takes prepositions such as 'in order to' ('He killed Smith in order to inherit his estate.') 'Motive' in the dispositional sense (3) usually takes a preposition such as 'from' ('He killed Smith from greed or avarice.') and ordinarily doesn't admit of dating, at least not from a fairly exact time as 'motive' in sense (2) usually does. (Which fact explains Ryle's insistence that motives in this sense aren't happenings and the, I think, mistaken inference he draws from this in denying that motives can be causes. Although this paper is not directly concerned with the problem of mental causation as such, I should be prepared to argue that both (2) and (3) can be properly considered as causes of action.) (3) When it figures in the explanation of an action does so more on the level of "character" or "character-traits" while (2) figures more on the level of particular decisions or choices made by an agent. In "the standard case" the occurrent sense of 'motive' occupies "the foreground" of explanation and the dispositional sense "the background".

If this brief sketch of the three principal senses of 'motive' is correct, it helps to explain what Miss Anscombe has noted where the wider and more diverse application of 'motive' over 'intention' is concerned. It seems that where the applications of 'motive' and 'intention' do overlap it will be when we have in mind 'motive' in the occurrent sense. 'Intention' ordinarily is concerned with decision or choice rather than with character or the disposition to act, and it admits of dating in the way that the occurrent sense of 'motive' does. That 'intention' is not interchangeable with 'motive' in the first sense of 'motive' noted above is obvious: if our police detective tells Jones that he (Jones) had the intention of murdering Smith this is far stronger than merely to report the presence of a reason (in this case a benefit to be derived from the murder); it is in short to make an accusation, to state that the accused actually did intend to commit murder. Also, while I am unable to give a comprehensive account of why 'intention' insofar as it is interchangeable with 'motive' is interchangeable with 'motive' in the occurrent sense, both 'intention' and 'motive' in the occurrent sense quite naturally take prepositions like 'in order to', and it is in such cases that the interchange of 'motive' and 'intention' works so well as to tempt us to overstate the resemblance between 'motive' and 'intention'.

---

[1] *The Concept of Mind* (London, 1949), pp. 83–115.

Perhaps we can now understand better why it is that an intention may be regarded as a forward-looking motive, provided it is 'motive' in the occurrent sense that we have in mind. An intention is a forward-looking motive (in the occurrent sense) in that it involves a decision or choice to bring about a certain, rather specific future state of affairs. Unlike 'motive' in the dispositional sense where the motive in question is usually quite general (ambition, vanity, curiosity, *etc.*) and not usually related to any one specific goal, 'motive' in the occurrent sense usually looks to the future with a definite goal in mind (for example: 'What was Nixon's motive in campaigning so hard for Goldwater in 1964?' 'He did it in order to improve his position within the Republican Party and to secure the nomination for himself in 1968.') Unlike 'motive' in the sense of there being a reason for acting, 'motive' in the occurrent sense involves an agent's decision or choice to act in such a way as he believes will bring about the attainment of a given goal. What I have said thus far about 'motive' in the occurrent sense holds true of 'intention' as well; in what respect then is it misleading to say that an intention is a forward-looking motive? The answer, I believe, is that while this is true it obscures a fundamental difference as to how intentions and motives in the occurrent sense manage to look forward.

While 'intention' and 'motive' will usually interchange easily when it is the goal or end of an action that we are speaking of, the interchange runs into difficulties when we wish to speak of means to that goal or end. The Pragmatist Within, impressed by the ease with which the interchange works when it is only the goal or end of action that we wish to characterize, may wish to deny altogether, or at least to minimize the importance of, the failure of this interchange when we complicate the story by introducing the means-end relationship. However, things are not so simple as he might wish, as can be shown by the following example.

Let us suppose that John's goal is to marry Wanda Lou, but her father, Farmer Brown, is so adamantly opposed to this marriage that John believes the only way to marry Wanda Lou is to kill her father, which he proceeds to do. If asked why John killed Farmer Brown, we can reply equally well either that his intention was to marry Wanda Lou or that his motive was to marry Wanda Lou. No difficulty confronts us until we begin to talk about the killing of Farmer Brown as a means to the end of marrying Wanda Lou. We can say that John's intention was to kill Farmer Brown and that his motive was to marry Wanda Lou, but we cannot say that John's motive was to kill Farmer Brown and that his intention was to marry Wanda Lou—unless, of course, it is a different story that we wish to tell in the second case. Among other things, the second way of speaking makes the death of Farmer Brown sound more like the goal or end John really had in mind, and his intention to marry Wanda Lou now seems less like an intention and more like a pretext or rationalization for the murder of her father.

This complication concerning the interchange of 'motive' and 'intention' persists if we extend the story by expanding upon the means-end sequence. If we do this by elaborating further upon the sequence of actions John followed in order to marry Wanda Lou (he purchased a gun, he practised firing it, *etc.*), these actions must still be spoken of in the language of intentions rather than in the language of motives ('Why did he purchase a gun?' 'His intention was to use the gun to kill Farmer Brown, but his motive was to marry Wanda Lou.') If we expand upon the means-end sequence by now making the marriage to Wanda Lou a means to some further end, say the acquisition of Farmer Brown's Bugatti Royale, then the marriage is no longer John's motive but (only) his intention.

What I have done in this paper is, of course, not offered as a comprehensive account of the resemblances and differences between 'intention' and 'motive'. It does, however, point out several important resemblances between 'intention' and 'motive' in the occurrent sense, as well as one crucial difference. Either 'intention' or 'motive' in the occurrent sense can be used to mark the goal or end of an action, but where the means-end relationship is concerned the interchange fails. Here 'intention' and 'motive' in the occurrent sense face the future in different ways, with 'intention' marking the means and 'motive' the end.

University of California, Santa Barbara

WHAT KNOWLEDGE IS NOT

By W. R. ABBOTT

PROFESSOR Edmund Gettier showed some time ago (ANALYSIS, Vol. 23 (1963), pp. 121–3) that the analysis of knowledge as justified true belief fails. The source of the difficulty he found lies in the notion of justification, which would permit someone to have a justified true belief that p and yet not satisfy us that he knows that p.

In this paper I offer an independent argument to show that knowledge cannot be justified true belief; a justified true belief may be necessary for knowledge, but the two are logically distinct. The logical distinction is that, when 'K' is read as 's knows that . . .', the schema

$$(1)\ (K(p) \& K(q)) \rightarrow K(p \& q)$$

is valid, whereas, when 'K' is read as 's has a justified true belief that . . .' (1) is not valid, *i.e.*, is not true for all substitution instances for p and q.

If someone, Smith, claims to know that there is someone at the front door, and also claims to know that the front door is locked, can he intelligibly maintain that he does not know that there is someone at the front door *and* it is locked? Surely not. If he does not know p and q, then the denial of $(p \& q)$ is compatible with all the propositions that he does know, and the revelation that $(p \& q)$ is false is (logically) compatible with all the propositions he knows. But if $(p \& q)$ is false, then that means either that p is false or q is, or both are. In that case he cannot know p and know q, since knowing p entails that not-p is incompatible with what he knows, and his knowing q entails that not-q is incompatible with what he knows; thus his knowing p and his knowing q entail his knowing $(p$ and $q)$.[1]

To say that someone s has a justified true belief is to say (a) that p obtains, (b) that s believes p, and (c) s is justified in believing p (has adequate evidence, *etc.*). Let 'Jp' be read as 's is justified in believing p', 'Bp' as 's believes p'; the claim that reading 'K' in (1) as 's has a justified true belief that . . . ' can then be rendered as:

$$(1^*)\ ((Jp \& p \& Bp) \& (Jq \& q \& Bq)) \rightarrow (J(p\&q) \& (p\&q) \& B(p\&q)$$

Since p and q jointly entail $(p \& q)$, that part of the consequent above gives no trouble for the validity of (1*). Presumably Bp and Bq jointly entailed B$(p \& q)$, but if they don't then the invalidity of (1*) is established. The invalidity of (1*) is more readily established, however, by the failure of Jp and Jq jointly to entail J(p & q). If someone is justified in believing p and is also justified *on independent grounds* in believing q, the degree of confirmation of the joint proposition (p & q) is less than the degree of confirmation of either of the original propositions. No matter what justification one has in predicting the outcomes of particular football matches, the justification one has for one's entry into the football pool falls well below the justification of any one of the predicted outcomes. As the number of matches is increased above two, the point is perhaps easier to see, but the logical point remains the same. Thus it may happen that the joint proposition, (p & q) is not well enough confirmed to meet the standards at hand. In fact 'is confirmed', 'has adequate evidence', and 'is justified' all fail to satisfy schema (1). The invalidity of (1*) follows from what has been shown and the fairly obvious fact that J(p & q) is not entailed by any of the other constituents of the antecendent in (1*).

---

[1] This argument has a formal counterpart in Hintikka's epistemic logic in *Knowledge and Belief*, Cornell, 1962.

It is a commonplace that when we are asked to show what is meant by the term 'of equal length', we produce objects which we show to be of indistinguishable length, and a superficial sort of verificationism would lead to the view that what we mean by 'of equal length' is the same as what we mean by 'of indistinguishable length'. It is equally a commonplace that the difference between these two phrases is shown in the transitivity of 'of equal length' which is no part of 'of indistinguishable length'. Clearly the two phrases are closely connected, but, equally clearly, no examination, however careful, of a case of equal lengths will reveal to the examiner how that differs from a case of indistinguishable length. Similar remarks apply to the difference between justified true belief and knowledge; and the conclusion to be drawn is that any analysis of knowledge which fails to attend to the logical features of 'knows that' is bound to fail. But that should be no more discouraging than being told that 'equal length' cannot be defined simply as 'indistinguishable length'; the point is that we have clarified a conceptual relationship.

University of Waterloo

NOTES

The Analysis Committee consists of: Chairman, P. T. Geach; Secretary, Miss M. A. Boden; J. H. Benson, Alec Fisher, A. Harrison, R. F. Holland, S. Körner, A. C. MacIntyre, A. M. MacIver, B. Mayo, R. G. Swinburne, A. R. White. This committee is responsible for appointing and advising the editor and for the general policy of the paper.

Subscriptions. The subscription to Analysis, Vol. 31, is £1.20 in Great Britain and £1.40 abroad (U.S.A. $4.00), post free. Each volume comprises six numbers, five of 32 pages and one of 48 pages, appearing within the academic year—in October, December, January, March, April and June. Orders should be sent to Basil Blackwell, 49 Broad Street, Oxford, or placed with any bookseller.

30 JUL 1971

Contributions. Articles submitted for publication should be addressed to Professor Peter Winch, Department of Philosophy, King's College, Strand, London, W.C.2. Contributors are asked to note the following requirements:

Articles should **normally** not exceed 3,000 words in length. **Occasionally**, however, longer contributions can be accepted;

They must be typewritten in double spacing on one side of the paper only;

Footnotes should be kept to a minimum and wherever possible avoided altogether;

Single quotation marks should normally be used, except for purposes of internal quotation and "scare" quotes.

Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

Contributors in the United Kingdom should enclose a stamped addressed envelope of suitable size; if immediate acknowledgment is required a stamped postcard should also be enclosed.

Overseas contributors who wish to have their MSS. returned should send an envelope **and international reply coupons** of the requisite value, whether for air or surface mail.

Proofs are normally read by the Editor. A contributor who wishes to read his own proofs should notify the Editor at the same time as the MS. is submitted.

Offprints. Contributors will be supplied free of charge with a limited number of offprints of their articles. Further copies may be ordered at a charge from the publishers. Orders for these, if required, should be placed **direct with the publishers** when an article has been accepted for publication.

The copyright of articles printed in Analysis remains the property of the author, but contributors are strongly advised, in their own interest, to consult the Editor before consenting to the reprinting of their articles.

PRINTED IN GREAT BRITAIN BY A. T. BROOME AND SON, 18 ST. CLEMENT'S, OXFORD

ol. 31 No. 5 (New Series No. 142) April 197

ANALYSIS

Edited by
Peter Winch

CALCUTTA UNIVERSITY
CENTRAL LIBRARY

CONTENTS

On near knowledge — James Cargile
Knowledge without belief — Carolyn Black
Williams and Stroud on Shoemaker's sceptic — William G. Lycan
Physiological discoveries: criteria or symptoms — Stewart Candlish
Three mistakes about retributivism — Jeffrie G. Murphy
On death as a limit — James Van Evra

BASIL BLACKWELL · BROAD STREET · OXFORD

Twentyfive pence

ANALYSIS 31.5 APRIL 1971

ON NEAR KNOWLEDGE

By JAMES CARGILE

CONSIDER the following premises:

1. If someone believes that p and is right, and his belief is of kind k, then he knows that p; and conversely, if he knows that p, then he has a belief of kind k that p.
2. If someone has a belief of kind k that p, then he may, on that basis, form a belief that the material disjunction, $p \vee q$, is true, which is also of kind k.
3. If someone does not have a true belief that p and does not believe that q, and (rightly) does not consider that if p were false, this fact would lend any support to the hypothesis that q, then he does not qualify as knowing that $p \vee q$.
4. Someone may have a false belief which is of kind k.

These premises are inconsistent. Premise 4 allows the possibility of a case in which someone has a belief of kind k that p, which is not a true belief, on the basis of which he may, by premise 2, form a belief that $p \vee q$, which is also of kind k, even though he does not believe that q. If we choose a case in which q is true, but the hypothesis that q is not supported by the denial of p, the agent will be qualified as knowing by premise 1, and disqualified by premise 3.

If we interpret 'belief of kind k' as 'justified belief', then the inconsistency of 1–4 is the basis for Edmund Gettier's[1] conclusion that knowledge is not justified true belief. This conclusion is the negation of premise 1, and is essentially derived by taking premises 2–4 as true on this interpretation.

This is a very sound point, understanding 'justified belief' as 'perfectly reasonable belief' or the like. Someone may perfectly reasonably believe something, be right, and yet not be counted as knowing it. However, when we say 'his confidence turned out to be justified (unjustified)', we are using 'justified' in a different way, which includes being right. So interpreted, premise 4 would be false, and the inconsistency of 1–4 would not serve as a basis for rejecting 1.

There is a further possible interpretation of the very loose term 'justified'. Philosophers who accept the formula 'knowledge is justified true belief' may mean merely that someone may believe something and be right and yet not be counted as knowing it, because his reasons (or other warrants) for believing aren't good enough. In this context, to say a true belief is 'justified' may mean just that it is based on reasons (or other warrants) which are good enough—good enough to qualify it as knowledge.

---

[1] 'Is True Justified Belief Knowledge?' ANALYSIS 1963, pp. 121–22.

It might be objected that this interpretation makes the formula a circular account of knowledge, that it becomes the triviality 'knowledge is true belief based on good enough reasons (or other warrants) to qualify as knowing'. This charge is fair, but that is only all the more reason to be dissatisfied with a refutation of the formula. Refuting a circular definition, showing it false, should not be easy, to say the least.

In what follows I will consider a reading of 'belief of kind k' which should make premise 1 trivially true. It is

> (A) a belief based on reasons (or other warrants) which are good enough, *for the given situation*, that, if it is a true belief, it qualifies as knowledge.

'Justified' seems to me a term loose enough to be read in this way (which is not to imply that A itself is very precise). But, when read in this way, premise 1 is trivially true.[1] We may then ask which of the remaining premises should be given up, since at least one of them must be given up for the sake of consistency.

Since premise 3 does not contain any reference to 'belief of kind k', it should be true on any interpretation of this phrase if it ever was true. Furthermore, it only needs to be true in one suitable case in order to make premises 1, 2 and 3 incompatible. And even if some particular case could be found in which someone qualifies as knowing p V q despite the failings described in premise 3, it seems clear that suitable cases in which 3 is true can be found.

This leaves premises 2 and 4. They are not strictly independent because given the additional premise,

> 5. If someone knows that p, then he knows that p V q, or at least will if the question (whether p V q) is put to him, and he has his wits about him,

premise 2 can be derived from the denial of 4. If there can be no beliefs of kind k which are false (the denial of 4), then any such belief will be knowledge (by 1). So (by 5) from any belief of kind k that p one can pass to knowledge that p V q. And (by 1) this means having a belief of kind k that p V q. So from 1, 4 and 5, one can pass from a belief of kind k that p to a similar belief that p V q, which is what premise 2 says.

Since 5 is surely true, as is 1 on reading A, this means that we cannot deny both 2 and 4 and, since we are also accepting 3, we cannot affirm both 2 and 4 either. We have to choose between them.

---

[1] Except for those who deny that knowing involves believing. This will be taken into account later.

My view is that 4 should be rejected. For what it is worth, if 4 is false, then 2 becomes obviously equivalent to the obvious 5, while 2's being false would not similarly make 4 obvious. But of course this is not enough to settle the question in favour of 4.

A notable example of a philosopher who would disagree with this proposal to reject 4 is Norman Malcolm. He says

> As philosophers we may be surprised to observe that it *can* be that the knowledge that p is true should differ from the belief that p is true *only* in the respect that in one case p is true and in the other false. But that is the fact.[1]

Of course Malcolm does not mean that the *only* difference is that in one case p is true and in the other p is false. Even if we take the broad use of 'proposition' in which a proposition can be true in one place or time and false in another, the time or place will still be different, and this is an additional difference over the difference in truth value for p in the two cases. Malcolm must mean just that there can be two situations which differ in only one respect relevant to the question of whether the belief involved in each situation is knowledge—that what is believed in the one case is true, in the other, false. And this entails the truth of premise 4 on interpretation A.

Strictly speaking, it might be wrong to characterize Malcolm as talking about a case in which a belief qualifies as knowledge at all. For he may be one of those philosophers who deny that when someone knows that p, he believes that p. One basis for this is the correctness of 'he doesn't (merely) believe, he knows', which seems to imply that knowing not only does not entail believing, but precludes it.

My first impulse, when faced with this difficulty, was to say that all I mean by 'believes' is 'regards it as true that' or 'accepts it as true that', or some such, where I could say that, surely knowing entails that much. But we can also say 'he doesn't *regard* it as true, he *knows* it is true', 'he doesn't *accept* it as true, he *knows* it is', and so on. So someone might say that knowing is just a fundamentally different matter from believing, and no rewordings will get around this gap.

In my opinion, all these examples in which we have 'he doesn't—he knows' turn on the use of emphasis to introduce the qualification 'merely', so that all we really have is that knowing is not *merely* believing, *merely* regarding it as true, and so on. After all, to say 'he doesn't believe it' or 'he doesn't regard it as true', on the grounds that he knows it, would ordinarily be grossly misleading if not accompanied by the explanation that this is your ground for saying it.

However, there is really no need to settle here the question whether knowing includes believing. For when I ask whether someone's belief that p qualifies as knowledge, I could just as well ask instead whether he is in a position such that he qualifies as knowing that p. If I spoke in this way, the basic point which I am defending, namely, that 4 is false, would only require a little rewording. Instead of 2, I would write

---

[1] *Knowledge and Certainty*, Prentice-Hall 1964, p. 60.

> 2′. If someone is in a situation such that he is possessed of qualifications which would rank him as knowing that p, subject only to the condition that p is true, then on that basis, he may go on to be in a similar position with respect to the material disjunction, p V q.

And instead of 4, I would write

> 4′. Someone may be in a situation such that the only reason why he does not qualify as knowing that p, is that p is false.

This 4′ is just what Malcolm claims. All that I have to say could be reworded so as to be concerned solely with denying 4′, without ever speaking in terms of beliefs qualifying as knowledge. For example, rather than saying 'Someone knows that p if his belief that p is true and is based on a means of assessing the situation reliable enough to qualify him as knowing', I could say instead 'Someone knows that p if p is true and he has assessed the situation by means reliable enough to qualify him as knowing, given that any other requirements for knowing have also been met'. I shall speak in terms of beliefs qualifying as knowledge, but nothing that is important here depends on this.

Malcolm gives an example of a pair of cases which he considers to exemplify his point. In each case we are talking about going to Cascadilla Gorge and, in each case, someone (let us call him Jones) says, 'I know that it won't be dry'. In each case, Jones can support this with the true observation 'I saw a lot of water flowing in the gorge when I passed it this morning'. But in one case the gorge is not dry, in the other it is. Of the former case, Malcolm says 'If we went and found water, there would be no hesitation at all in saying that you knew'. Of the latter, he says that everything happens as before, 'except that upon going to the gorge we find it dry. We should not say that you knew, but that you *believed* that there would be water.'

Now it is certainly true that in the latter case Jones did not know, but merely believed, that the gorge is not dry. But it does not follow that he only missed knowing by the mere fact that he is wrong, and that the case in which he knows differs *only* in that, in that case, the water is there, while in the other case it isn't. A lot of water remains to be accounted for. Whatever became of that flood in the gorge that was observed in the morning, that it should not even be wet by afternoon? This is not something totally inexplicable. Even if it is a freak of nature, it will have some cause. And even if it is perfectly reasonable for Jones to have overlooked or underestimated this causal factor, even if it would have been impossible for him to have known about it, the fact remains that appreciating the significance of this factor was necessary in this case, for predicting the condition of the gorge. So even if Jones had *good* reasons for thinking the gorge would be wet, they were not *good enough* to let him correctly assess the situation. And in my opinion, to call them good enough to have qualified Jones as knowing, if only he hadn't been wrong, is quite implausible.

Ramsey says, 'I have always said that a belief was knowledge if it was (i) true, (ii) certain, (iii) obtained by a reliable process'.[1] Unfortunately, as Ramsey himself continues, 'The word "process" is very unsatisfactory'. We may in some cases say someone arrived at his belief by a process, say autosuggestion, but in many cases, the word 'process' just won't be applicable. I believe I am now writing a paper—but by a process?

However, I think we can speak, if not of 'a reliable process', of 'a reliable means of assessing the situation'. Reliable means of assessing the situation may provide those "other warrants" mentioned in reading A. When I see that something is before me, I don't have reasons to offer for thinking so, but my belief is nonetheless based on a reliable means of assessing the situation and is thus warranted.

In these terms, my disagreement with Malcolm can be put as follows: it is not true that someone can have a false belief based on means of assessing the situation which are reliable enough to qualify the belief as knowledge, if only it were true.

If we consider cases in which the only relevant difference really is that you are right in one and wrong in the other, we will find that the proper conclusion will be that for that very reason, you did not know in either case. For example, suppose Jones is standing on the corner, watching all the girls go by. Five people pass and he forms the five firm convictions 'there goes a woman—there goes a woman—and there goes another—and another—and another'. Actually, it is a group from a nearby convention of female impersonators. But the third member of the group is the group's secretary, a female. Now the evidence for each of these five beliefs does seem the same in all respects relevant to assessing Jones's title to knowledge, and the beliefs are also similar in the relevant respects. But the evidence isn't good enough to save Jones from error in four of the cases *in the same situation*, and that is why no one would credit Jones with knowing the third person to pass by is a female, even though he is right.

Meanwhile, someone else in another city, Smith, may be standing on a corner, watching all the girls go by, and he will certainly be credited with knowing when a woman is passing (even though there are many difficult cases nowadays, there are still lots of obvious cases where the sex of a passerby can be told at a glance, and we may focus on those cases), even though his evidence is in one clear sense the same as that evidence which failed to provide Jones' title to knowledge, even when he was right. It is different, though, in one relevant respect. It is set in a situation in which that sort of evidence is adequate, and Jones' evidence is not.

---

[1] F. P. Ramsey, in *Foundations of Mathematics*, Routledge and Kegan Paul 1960, p. 25.

These two objections to Malcolm's claim can be summarized by schematizing his claim. Malcolm claims we can have two cases as follows:

Case 1. belief+evidence+truth=knowledge
Case 2. belief+evidence+falsity ≠ knowledge

That is, the beliefs and the evidence are congruent, and only the difference in truth makes the difference as to the question of knowledge. My objection is that (1) when the one case really does involve knowledge and the other doesn't, the evidence will not really be congruent as far as the question of knowledge is concerned, and (2) when the beliefs and evidence really are congruent, and there is falsity in one case, then there is not knowledge in either.

Malcolm did not expressly deny premise 4 on reading A; he said something which entails denying it. One reason he might have had for saying what he did could be accepting premise 1 in terms of the following reading for 'belief of kind k':

> (B) a belief based on reasons which are good enough, *for any situation*, that, if it is a true belief, it qualifies as knowledge.

Accepting premise 1 on this reading would have the consequence that if a certain kind of evidence ever proved inadequate to protect you from error, then it never could in any situation be good enough to qualify you as knowing. That is, it would have this consequence *unless* the evidence is completely exonerated from blame in the case where it allowed error. And this consequence is a profoundly sceptical one, which Malcolm would understandably wish to avoid. But avoiding it by saying that, surprisingly enough, the *only* difference between the case where there was error and the case where you were right which is relevant to the question of knowledge is the difference in truth value of what is believed, is a bad way to avoid this scepticism. It has the consequence that Jones did know the third passerby was a woman, which is a false consequence. It is better to reject premise 1 on reading B, and to accept it on reading A instead, and then to reject premise 4.

To this it might be replied that my claim (2) is based on only one favourable case, in which one true belief is disqualified as knowledge because of four similar beliefs being false in the same situation. What if we have a situation in which there are thousands of true beliefs and only one false one, and my rule requires us to disqualify all the true beliefs from being knowledge just because of the false one?

There could indeed be such cases, but in my opinion, my rule applies to them. For one example, suppose that Jones has a job requiring him to sit by a conveyor belt carrying boxes of watermelons and to mark on the side of each box whether it contains 5, 6, or 7 melons and then close it. He marks a thousand boxes a day and tends to make a mistake about once a week. Does he not know that this closed box he has just marked contains 7 melons, in spite of the fact that 100 boxes ago he mistakenly marked a box with 6 melons in it as having 7, and falsely believed it to contain 7?

I would say that he does not know. Given a closed box he has marked with a '7', Jones would, in my opinion, be speaking falsely if he said 'I know this box contains 7 melons'.[1] For he would have to admit that he does in fact make mistakes in marking occasionally, and he has no reason to think that this case is not one of the mistakes, other than an appeal to probability. His assessment of the number of melons is by the very same means which occasionally lets him make a mistake. If Jones had paid special attention to this box, or could show that mistakes only occur when he is tired, and that he was not tired when this box was checked, it would be a different matter.

Of course, if Jones is right about the box in question it is not merely by chance. It is in large part because he has assessed the box by reliable means. But it was not by means reliable enough to rule out all probability of error, and thus was not by means reliable enough to qualify him as knowing.

Here it is important to stress that, in my opinion, it would be *very* easy for Jones to make absolutely certain that there was no probability whatsoever of his being wrong about the box in question. He could open it and do a careful recount. I do not hold with the view according to which zero probability applies only to what is logically impossible.

Perhaps it is worth considering another case similar to the preceding one, which might be advanced in defence of Malcolm's claim. Suppose that an unbiased coin is going to be tossed 1,000 times by a machine. Don't we know perfectly well it won't land heads every time? If it were a question of staking anything on the outcome, couldn't we say 'We *know* it won't land heads every time'? But then, if this machine run of 1,000 tosses were performed $2^{10,000}$ times, it would be very likely that one such test would yield a run of 1,000 heads. And if someone had said, before this test, 'I *know* it won't be all heads' wouldn't he be an example of someone who would have known, if only he had not been wrong? In this case, there are no causal factors he needed to be aware of, but wasn't. Pure chance happened against him.

---

[1] It isn't just that Jones' *saying* 'I know' would be wrong. I know that I have two hands, even though saying this may be generally out of place. It is better to view the question whether someone knows as if you were on a panel drawn up to decide whether various cases qualify as knowledge. This will allow a little freedom from the constraints of everyday discourse.

Here I can only record a difference of opinion. Suppose that $2^{10,000}$ people are each confronted with the prospect of our tossing machine's throwing the coin 1,000 times. Then I concede that it is natural to allow that each such person would be reasonable to suppose that his series of tosses will not all be heads. But to say that all those who turn out to be right, *knew*, while those who are wrong didn't, *merely* because they were wrong, seems absurd to me. We would be saying 'It is extremely likely that some of these people will be wrong, but those that aren't wrong, that turn out to be right, *knew* that was how it would turn out'. This would support Malcolm's claim. I can only say that it sounds wrong to me.

One complication in this case, of course, is that there is always the question whether the coin-tossing machine has been unbiased throughout the trials. But since that is only a complication in setting up a case to oppose my view, it is of no concern to me.

University of Virginia

KNOWLEDGE WITHOUT BELIEF*

By CAROLYN BLACK

IS one of the criteria for knowledge belief? Does knowledge entail, include, involve, or necessitate belief? Must we include in a description or definition of knowledge a belief condition? While many philosophers assume or stipulate that a definition or description of knowledge is in part dependent upon belief, I do not think that this has been demonstrated, or even convincingly argued for. This paper is about some of the reasons why. My feeling is that most knowledge clearly does not include belief and that the rest is not elucidated much, if at all, by describing it as comprising belief.

* I made some of these points in my Ph.D. thesis, *Belief*, Bryn Mawr College, 1967.

Let us begin with a question Moore asks in *The Commonplace Book* 1919–1923 (1962, p. 115):

> There certainly is *a* common use of belief in which 'I believe' entails 'I don't know for certain'. Is there another in which 'I know for certain' entails 'I believe'? One reason why it seems so is because 'I thought I knew' entails 'I believed'.

Is there a kind of belief in which 'I know for certain' entails 'I believe'? or: Is there a use of 'believe' or 'belief' which is indicated or implied by 'I know'? Let me summarize my view with respect to these questions. Later I will argue for the main points. When in good faith I make the statement, *e.g.*, 'I know that the cat is on the table' it is true that: (1) I *may* truly be said to believe (think) and not know that the cat is on the table if, say, the cat is not on the table; (2) I would not claim that I merely believe that the cat is on the table; and (3) there may be a dispositional belief here. *I.e.*, I may, if asked, and *if* a sensible setting for the question could be provided, say that I believe that I know that the cat is on the table. Instead of using the term 'believe' here I may use the term 'know', depending on how careful I want to be about my original claim and how sure I am about whatever it is that I claim to know. The theoretically possible infinite regress concerning believing or knowing that one knows does not seem to be vicious in these cases. It is rarely of any concern and quite introspectively difficult, if possible, to say sincerely, *e.g.*, that one believes that one believes that one knows that the cat is on the table.

Similarly, in an epistemic claim which does not include the term 'know', *e.g.*, 'The cat is on the table', it is true that: (1′) As in (1), I may truly be said to believe that the cat is on the table; (2′) as in (2), when I am in a position to say, 'The cat is on the table' I probably would not say that I believe that the cat is on the table; (3′) there may be a dispositional belief here, with the same qualifications as in (3). I may say that I believe that I know that the cat is on the table or that I just believe that the cat is on the table. However, the fact that the claim, 'The cat is on the table' may be followed by the question, 'Do you know that, or only believe it?' shows that at least one kind of belief is not a condition for purported statements of fact.

Let us look at some of these items further. Utterances of the kind '*p*' and 'I know that *p*' do not conventionally require utterances of the kind 'I believe that *p*'. One does not claim that one believes that the cat is on the table when one is prepared to claim that the cat is on the table, or that one knows that the cat is on the table. This point may be viewed as part of the more general and familiar one (due to Grice, I believe), namely, it is a convention of language that in many circumstances we make the strongest epistemic claim of which we are capable.

E.g., we do not say that we are going to Mexico or Canada when we are going to Mexico. That we are going to Mexico or Canada when we are going to Mexico is true and, if most logicians are right, entailed by our going to Mexico. But we seldom want to claim this usually irrelevant truth.

The utterance 'I believe' aside, what does it mean to say that the utterance 'I know' implies that I believe? This might seem to follow if knowledge comprises belief. Does it mean that I dispositionally think (believe), *i.e.*, that I would if asked (and when could I be asked? In the philosophy class?), say that I believe *that I know*, or that I am not lying when I say 'I know' or that the true or sincere utterance 'I know' means that I at least think? There are many alternatives here, some of which we will explore further later. We can settle one point now. If you do not know that I know and say that I think (believe), you could be wrong. Cases of reminding are decisive. If you find out that I recently learned that *p* and have momentarily forgotten it, you would be wrong in saying I believe that *p* but not in saying that I know it. You could not say, 'She believes but she does not believe that she believes' but you might say, 'She knows, but she does not believe or know that she knows'. Radford gives more elaborate examples which support this point (ANALYSIS, 1966).

The treatment by Moore of his question lends implicit support to the view that knowledge does not comprise belief. Let us look at the details of his remarks. He notes that the dictionaries of both Johnson and Webster indicate that there are two senses of the word 'belief'. One sense of 'belief' is compatible with proved knowledge, Moore says, while the other sense of 'belief' is incompatible with it. He expands

> Johnson defines 'believe': 'To credit . . . from some other reason than our personal knowledge', which says that 'S believes *p*' entails 'S does *not* personally know *p*'. But he defines 'credit' 'to believe'; and if you substitute 'believe' for 'credit' in the first definition, it follows that 'S personally knows *p*' is *not* incompatible with 'S believes *p*'. Webster recognizes the incompatible one: 'to be persuaded of the truth of, upon evidence furnished by . . . circumstances other than personal knowledge'; but he also recognizes the other one by defining 'persuade' 'cause to believe' (*Commonplace Book*, p. 115).

Notice that 'belief' in these dictionaries is explicitly granted a sense in which it is incompatible with knowledge and only implicitly and unintentionally a sense in which it is compatible.

Moore does not answer his question as to whether there is a sense of 'I know for certain' which entails 'I believe'. 'I thought I knew', which can entail that I believed, is quite different, so it should not be a reason to answer the question affirmatively. The most obvious way in which 'I thought I knew' entails that I believed is that 'thought' and 'believed' overlap in the sense that they can be synonyms. There are plenty of examples of this in the present paper. There is another, perhaps more interesting, way in which there may be an entailment. What Moore must be noting is that 'I thought I knew that *p*' amounts to 'I believed that *p*'. There may be cases of which this is true, but there are others of which it is not. For 'I thought I knew that *p*' may mean that I thought (believed) that *p* was the case, while 'I believed that *p*' may mean that I believed *p* merely probable. 'I thought I knew the human cell has 48 chromosomes, I thought it was true that the human cell has 48 chromosomes' may not mean that I merely believed or thought it probable that the cell has 48 chromosomes. A claim of certainty, when wrong, does not reduce to a claim of probability. With knowledge which is on the surface non-propositional, a more decisive case can be made. Consider: 'I thought I knew the names of all the states (but I got them wrong when I tried to recite them)'. 'I believed the names of all the states' is meaningless.

Some argue that while 'I thought I knew that *p*' might not easily reduce to 'I believed that *p*', if he thought that he knew that *p* when he did not, I can truly say that he believed that *p*. It seems to me that the most obvious thing that we can truly say here is 'He thought he knew'.

In any case, sooner or later one must be concerned with the present tense. Much that is true of one tense is not true of another. While 'I believe the book is in the other room' may mean 'The book is in the other room', 'I believed the book was in the other room' does not usually mean 'The book was in the other room'. 'I think I know' has at least two meanings, neither of which lends support to the view that it includes that I believe. 'I think I know' might be a mere expansion of the statement 'I know' much as 'It is true that *p*' is sometimes a superfluous expansion of '*p*'. One usually does not claim that *p* unless one feels that it is true that *p*; one usually does not claim that one knows unless one thinks one does. Here, for 'I believe I know that *p*' to be an honest claim, I must think *p* certain. 'I believe that *p*', by contrast, may be an epistemically second best claim, one made when I think *p* likely but uncertain. This is not a merely linguistic point. For believing that *p* is characteristically weaker epistemically than believing that one knows that *p*. Knowledge that *p* does not entail belief that *p* in cases like this. Alternatively, 'I think I know' may be a qualification, an epistemically weaker statement than 'I know'. ('Well, I think I know that π is transcendental.') 'I think I know that *p*' does not entail 'I believe that *p*' in cases such as this either though these cases are perhaps epistemically closer to knowledge claims which supposedly include beliefs than the previous ones. They are also more easily confused with them.

The clearest cases in which knowledge positively excludes belief are ones like the following: (1) I say that my books are in my office. You ask: 'Do you believe that your books are in your office?' I say: 'No. I know that my books are in my office'. Suppose instead that my initial statement is: 'I know that my books are in my office'. Suppose you then ask: 'Do you believe that your books are in your office?' This is one of those questions where the first mistake is to ask it, and the second is to try to answer it.

(2) Moore asked if we believe in the existence of particular bodies, *e.g.*, of one's body and one's pen. One reason for saying no, he says, is that one knows that they exist (*Commonplace Book*, p. 121).

Notice how in these first-person claims, knowledge positively excludes belief. We deny belief when we claim knowledge, if the issue meaningfully arises. The exclusion is third-person as well. In these cases one may not ascribe belief to the claimant without his denying it.

It is useful to look at first- and third-person cases further in relation to the exclusion of belief by knowledge. Let us contrast the following questions: If I can truly say that A knows that *p*, can this include that he believes that *p*? and: If I honestly claim that I know that *p*, can this include that I believe that *p*? These two questions are importantly different, for we can correctly ascribe knowledge to A when he is not prepared to claim to have it. *E.g.*, we can remind him of a fact that he has forgotten, when, therefore, he need not even believe the fact. There is no first-person parallel. Our second question is perhaps more manageable if we set aside certain verbal issues. Let us ask: If I know, do I believe? While there *may* be no superior mental act involved in knowing and not in believing, nevertheless I may think that I know that *p* or by contrast think that I believe it. Either I think that I am sure or I think that I am not. The former is what interests us here. No one has shown that my believing truly that I know that *p* is my believing that *p*. In all the cases I can think of, when I actually know that *p*, I do not and would not claim to believe that *p*. One might be led to think otherwise because "what goes on in my mind" can, if I am not being fussy about whether I think *p* probable or certain, be the same as what goes on when I think that I know that *p* when I *do not.* But here we do not have knowledge including belief for we do not have knowledge. This it seems to me is the heart of the matter.

There are objections to the view that I am advocating which must be dealt with. Suppose I say, 'I know that the cat is on the table. I see him there through the window. I do not believe it.' Some philosophers say that what we mean is that we do not merely believe, we also know. I have seen no decisive argument for this. Keith Lehrer argues (*Philosophical Review*, 1968) that when we say, 'That's not a house, it's a mansion' we mean 'That's not *merely* a house, it's *also* a mansion'. He says this case is exactly parallel to saying 'I don't merely believe it, I also know it'. To the extent that the cases are parallel, my view is just as plausible.

Why call a mansion a house, once you've called it a mansion? Is a philosophy-psychology major a philosophy major? My inclination is to say no. But perhaps the best thing to say is not 'yes' or 'no' but simply, 'A philosophy-psychology major is a philosophy-psychology major'.

No so-called parallel, no matter how good, will be decisive anyway. As Austin says, we do not know by analogy, we only argue by it. Lehrer does not rest his case on the analogy. He thinks he has shown that knowledge *formally* entails belief. Here is his "proof":

(1) If S does not believe that p, then S does not believe that he knows that p.

(2) If S does not believe that he knows that p, then, even though S correctly says that p and knows that he has said that p, S does not know that he correctly says that p.

(3) If, even though S correctly says that p and knows he has said that p, S does not know that he correctly says that p, then S does not know that p.

(4) If S does not believe that p, then S does not know that p.

Lehrer's thesis that knowledge entails belief is supposed to follow by contraposition thus $((-B \supset -K) \supset (K \supset B))$. I do not believe that Lehrer has shown that $(-B \supset -K)$. His argument is a compendium of errors and assumptions. I have cited cases which make premise (1) false (p. 155f above). Some situations in which I do not believe that p are situations in which I think p not probable, but the case, situations in which I believe that I know that p, and not in which I believe that I believe that p. Premise (3) is questionable as well. If I were a Greek teacher and an unconfident student told me he did not believe that he knew the Greek alphabet, but recited it perfectly in an exam, I would say that he knew it despite his claim. *Pace* Lehrer, if one correctly says that p, and knows he has said that p, and still does not know that he correctly says that p, he may still know that p. Lehrer has not shown that if you do not have belief you cannot have knowledge.

If behaviouristic philosophers have shown anything, they have shown that a good test of knowledge, sometimes the sufficient condition of knowledge, is doing something. You know, and show me and yourself that you know how to swim if you jump into the ocean and start swimming whether either of us believed it previously or not. You know, and show me that you know that the cat is on the table when you claim this knowledge, take me to the table, and say 'Here. See for yourself'. Naturally, one may know something without doing anything, but remember where the proofs of puddings lie.

To conclude. When attempting to define knowledge, or when noting the necessary and sufficient conditions for knowledge, one does not have to include a belief condition. Indeed, all attempts along these lines have failed, as Gettier and others who have elaborated upon his examples have shown. ('Is Justified True Belief Knowledge?' ANALYSIS, 1963; Gilbert Harman, 'Lehrer on Knowledge', *Journal of Philosophy*, 1966.) Even Plato realized that justified true belief is not knowledge. But Theaetetus worries more about the *logos* than about the belief.

In first-person knowledge claims, one does not claim to have belief when one is ready to claim to have knowledge. There is no point in claiming to have both.

In third-person ascriptions of knowledge, as in first-person claims, it is epistemically unconventional to ascribe belief to someone to whom we are prepared to ascribe knowledge. Further, we may truly ascribe knowledge to someone when he does not have any positive beliefs about the matter.

While no one has shown that believing that one knows that *p* can include believing that *p*, there are many cases in which it cannot. A crucial epistemic ambiguity regarding the certainty or alternatively the probability of items of belief tends to lead one astray here.

Why have philosophers tried to explain knowledge in terms of belief? It may be an example of that craving for generality that Wittgenstein warns us of. It is easier, even if not accurate, to try to explain knowledge, in part, by means of a common concept which is already delineated to some extent, than to try to explain it without the use of that concept. It is easiest to lump together all the concepts one can. Similarly, it is easier to describe tangerines as orange-like, or *as oranges* plus a few minor differences, than to analyse or explain tangerines without bringing in their likenesses to other fruit. As Wittgenstein says, we must get over that craving for generality. Perhaps we should try isolation techniques for a change, as in biology.

Are there any good philosophical reasons for conflating knowledge and belief? I suggest that there are some reasons that will do quite nicely for the sceptic, but not for anyone else. If we never have knowledge, a big if, it is reasonable to suppose that what we have when we think we have knowledge is belief. But epistemological scepticism is another topic for another paper.

San Jose State College

WILLIAMS AND STROUD ON SHOEMAKER'S SCEPTIC

By William G. Lycan

I WANT to defend an argument of Shoemaker's, given on pp. 168–170 of *Self-Knowledge and Self-Identity* (Ithaca: Cornell University Press, 1963), against objections raised by B. A. O. Williams ('Knowledge and Meaning in the Philosophy of Mind', *Philosophical Review*, Vol. LXXVII (1968), pp. 216–228) and by Barry Stroud ('Transcendental Arguments', *Journal of Philosophy*, Vol. LXV (1968), pp. 241–256). I am not myself convinced that Shoemaker's argument is sound, but I hope to show that, if it is wrong, it is so for reasons other than those I discuss here.

1. I shall accept Williams' formulation of the argument:

> (1) Barring slips of the tongue, one who understands 'I am in pain' cannot utter these words with the intention of making a true assertion unless he is in pain. Therefore
> (2) If X understands 'I am in pain', then, if X says, 'I am in pain', intending to make a true assertion, X is in pain. Therefore
> (3) If it is possible to know that X understands 'I am in pain', it is possible to know that X is in pain. But
> (4) If 'I am in pain' has an established meaning, then it is possible for one language-user to know that another language-user understands it.

Let 'S' = 'It is impossible for one person to know of another that he is in pain'. Then

> (5) If S, then 'I am in pain' does not have an established meaning (3, 4). But
> (6) If 'pain' has an established meaning, 'I am in pain' has an established meaning; and only if 'pain' has an established meaning does 'S' have a meaning, since 'S' uses and does not just mention the expression 'pain'. So
> (7) If a true, and hence intelligible, assertion is made by the utterance of 'S', 'I am in pain' has an established meaning (6). Therefore
> (8) Not S (or rather 'there is a logical absurdity in uttering "S" intending to make a true statement') (5, 7).

The "logical absurdity" alleged in (8) is actually this: if S is intelligible, S is false (alternatively, S is either unintelligible or false).

2. Playing devil's advocate, Williams first doubts (p. 222) that '. . . the minimum sense of "understand" necessary to make (4) true is sufficient to make (3) true'. The sceptic might agree to (4) on a 'less ambitious' reading of 'understand', and agree to (3) on a 'stronger' reading of 'understand', and yet deny (3) on the less ambitious reading.

Williams does not say exactly what two readings or senses these are. Possibly he means something like this:

> '*X* understands (weak sense) "I am in pain" ' =df '*X* is disposed to utter the various words and to engage in the physical behaviour that the non-sceptical plain man considers to be appropriate to what he ordinarily calls "pain" situations'.

That is, *X* understands (weak sense) 'I am in pain' if he utters it usually when damaged and while wincing, *etc.*, and if, when others utter it when damaged and wincing, he treats them as if they were in pain. In this weak sense, the sceptic could say, (3) is false; the antecedent is admittedly true, but the denial of the consequent is just *S*, the sceptic's original claim. So the weak version of (3) begs the question against the sceptic.

Shoemaker has not merely asserted (3), however, but claims to have derived it from (1) and (2). It is to this derivation that we must look, then, in evaluating Williams' criticism and hunting down the putative *petitio*. Williams himself goes on to do so, saying that he will put his original objection 'in another way'.

3. He assails the move from (2) to (3). The inference, he says, has the form, 'If *P*, then *Q*; so if it is possible to know that *P*, it is possible to know that *Q*'. (Presumably, Shoemaker's value of '*P*' is, '*X* understands "I am in pain" and *X* says, "I am in pain", intending to make a true statement', and his value of '*Q*' is '*X* is in pain'.) Williams gives what he thinks is a counterexample to this inference-schema:

> . . . suppose that there is just one man, Robinson, who is fiendish and clever enough to commit a certain type of crime. Then it will be true that if a crime of that type is committed, it is committed by Robinson; but it does not follow from that that if it is possible to know that such a crime has been committed, it is possible to know that it has been committed by Robinson (he may make it impossible) (p. 223).

But this example is beside the point. Robinson makes it *practically* impossible for anyone to know that he has committed the crime, but he could not have made it *logically* impossible (imagine—the logically perfect crime!). And what we are concerned with is (*vide* Shoemaker, p. 169) the logical possibility of knowing that *X* is in pain.

Williams goes on to claim, correctly, that the aforementioned inference-schema as stated lacks an essential premise, 'It is possible to know that if *P* then *Q*'. So, presumably, if Shoemaker's argument is valid, it contains the suppressed premise, 'It is possible to know that (2) is true'. Shoemaker would grant this, since he 'represents (2) as being a necessary truth'. But, Williams says, it is open to the sceptic to

> wonder whether he *knew* it to be true at all. The very force of the argument might show him that to claim to know this came excessively close to claiming to know the sort of thing which he is claiming not to know (p. 223).

Perhaps so. But Williams seems to have shifted his ground: he now takes the concealed premise to be, 'The sceptic knows that (2) is true'—and this premise would be unnecessarily strong. All Shoemaker needs, by Williams' earlier account, is the *possibility* of (2)'s being known, and Williams has shown us no reason not to grant him this. (There is no suggestion on anyone's part that (2) is incoherent or otherwise logically unknowable; most sceptics, I think, would even affirm (2) if taken off their guard.) For all Williams has shown, then, Shoemaker's conclusion is successfully obtained: if the sceptic's claim is intelligible, it is false.

4. Stroud (p. 249) does not deny that Shoemaker has shown this to be so (that the sceptic's claim is either unintelligible or false). But it does not follow from this result, Stroud says, that the sceptic's claim *S* is necessarily false. Stroud regards this latter contention as being a 'further claim' of Shoemaker's 'which appears to be mistaken'; Stroud claims that '. . . it does not follow from the necessity of the conditional "if the sceptic's statement makes sense, then it is false" that the sceptic's statement is a necessary falsehood' (p. 249), and he reiterates this in another connection on p. 250.

But he gives no argument to block the entailment; and it seems reasonable, on whatever intuitive modal semantics we have, to think that the entailment does hold. For if, *necessarily*, p is either unintelligible or false, then it is impossible that p should be true. And if it is impossible that p should be true, then p is necessarily false, at least in a two-valued logic. My argument here can be tightened up by appeal to a "possible-worlds" semantics for necessity, reading 'unintelligible' as 'necessarily false': If it is necessary that p is either unintelligible or false, then, for any possible world W_i whatever, either p is unintelligible in W_i or p is false in W_i. Thus, there is no possible world in which p is true, and this suffices to show that p is necessarily false.[1]

---

[1] $LCALNpNpLNp$, incidentally, is a theorem both of S 5 and of T. I owe the proof of this to George Schumm.

It may be complained at the outset that 'unintelligible' should not be glossed as 'necessarily false'. To do so is indeed hasty, and may offend those who think that an ill-formed (and hence unintelligible) expression cannot be false at all. But my argument still shows that, if *S* necessarily is either unintelligible or false, then *S* is necessarily *untrue*. And that result, if correct, should be enough to alleviate worries brought on by the arguments of Shoemaker's sceptic.

University of Chicago and The Ohio State University

PHYSIOLOGICAL DISCOVERIES: CRITERIA OR SYMPTOMS?

By STEWART CANDLISH

TWO philosophers have in recent years made use of an argument which both proves too much and is ill-founded. Norman Malcolm, after exposing some genuine confusions on the part of physiologists attempting to investigate dreaming, says:

> If rapid eye movements during sleep became the criterion of dreaming, one consequence is that if someone were to tell a dream it could turn out that his impression that he dreamt was *mistaken*—and not in the sense that the incidents he related had really occurred and so his impression was not of a dream but of reality. The new concept would allow him to be mistaken in saying he had a dream even if his impression that he had seen and done various things was false. Another consequence is that it would be possible to discover that a man's assertion that he had slept a dreamless sleep was in error: . . .
> If the physiological criterion were adopted, telling a dream would only be a more or less reliable indication of dreaming. It would not be, as now, a matter of definition that someone who told a dream had dreamt. . . .
> Considering the radical conceptual changes that the adoption of a physiological criterion would entail, it is evident that a new concept would have been created that only remotely resembled the old one. . . .
> The desire to know more about dreaming should not lead scientists into transforming the concept in such a way that their subsequent discoveries do not pertain to *dreaming* (*Dreaming*, pp. 80–2).

Kurt Baier, arguing against the materialism of J. J. C. Smart, says

> . . . suppose we have the best possible physiological evidence to the effect that no one can have a pain in whose brain the process xyz is not going on. However good the evidence may be, such a physiological theory can never be used to show to the sufferer that he was mistaken in thinking

> that he had a pain, for such a mistake is inconceivable. The sufferer's epistemological authority must therefore be better than the best physiological theory can ever be. Physiology can therefore never provide a person with more than *evidence* that someone else is having an experience of one sort or another. It can never lay down independent and overriding *criteria* for saying that someone is having an experience of a certain sort. Talk about brain processes therefore must be about something other than talk about experiences (*The Australasian Journal of Philosophy*, Vol. 40, 1962, pp. 64/5).

There is a suggestion of the same argument in an article by R. C. Coburn (*The Journal of Philosophy*, Vol. 60, 1963, p. 91).

The argument even as presented shows clearly the absurdity of regarding physiological discoveries as a basis for avoiding the "problem of other minds" (V. C. Chappell has actually tried to do this on p. 19 of the Introduction to his collection *The Philosophy of Mind*), but it goes beyond epistemology to ontological issues. I wish to discuss it without raising the general question of the relationships among criteria, concepts and phenomena, by applying it to another related area in order to demonstrate its unfortunate consequences. The new application will not encounter the special complications of "privileged access", but that is not to say that it thereby misses the point; as I shall show, the possibilities of absurdity on which it relies outside the original area of its use are of an even grosser kind.

A few years ago I heard a physician point out in a radio discussion that, on death, all electrical activity of the brain ceases, and that in certain special circumstances—for instance, when there is cardiac arrest and there is some difficulty in determining whether a patient is dead or not—the cessation of brain waves for a certain period of time is used as the criterion of death. Now the use of the new EEG criterion brings with it the possibility of its conflict with the use of the established criteria, which are

(a) Absent pulse—Heart
(b) Absent respiration—Lungs
(c) Pupils dilated and fixed—Nervous system.

It could be said here that the EEG criterion is used only when there is doubt about the normal criteria, and so the possibility of conflict does not arise. This point, however, if it shows anything at all, shows only that the EEG criterion of death is superior to the ordinary criteria in that it deals with cases which they will not accommodate; and there is still the possibility that the EEG test for cessation of brain activity may be instituted as a routine test for death in, say, cases where there is not obvious gross physical damage.

What sort of conflict of criteria is possible once the EEG is used as a criterion of death? Well, it is possible that there should be no signs of brain activity from a man who exhibits normal signs of being alive like breathing, reading, feeling his heart quicken as he encounters the glance of his beloved, and so on; and it is possible that these signs should be absent, that he appears dead to us, while brain activity continues. The second case poses no particular problem for it is precisely the kind of difficulty that the introduction of the EEG criterion for determining death is designed to relieve; we would not send to the morgue a man whose brain still exhibited electrical activity, for such activity is a sign that the normal phenomena of life may return. (And surely our very reluctance to act thus suggests, despite the argument under discussion, that we are inclined to regard the EEG reading as more than mere *evidence* of life.) It is rather the first case which proves difficult, for it seems to show that we might all be mistaken about whether a man was alive or not, that we might all take him to be alive and yet a physiologist could demonstrate to us (and to him) his death; that he could be bundled, kicking and protesting, into his grave. Such a possibility is not only morally repugnant but apparently logically peculiar in that it is absurd to suppose that we might all be genuinely mistaken in this way; absurdity at least on a par with that involved in supposing a man to be honestly mistaken as to whether or not he is in pain. In fact, however, it is worse than that, for while in the latter case we could if we wanted to preserve the new criterion suppose, as Rorty suggests,[1] that one man is using the word 'pain' sincerely and incorrectly, and thereby posing unreal problems for physiologists, it is very odd indeed to suppose that we *all* used the word 'dead' incorrectly.

By applying in the last paragraph the argument under consideration, it seems to have been proved that cessation of brain activity cannot generally be used as a criterion of death; yet it *is* so used in certain circumstances, and it is quite feasible that these circumstances should be less restricted than they are at present. Are clinicians just wrong about this? There is a good case for saying that they are not; and it is worth stating what it is.

First, it is of the utmost importance that the possibility of this conflict of criteria, with its attendant absurdity, is never realised. Ruling out the employment of the EEG criterion of death requires more than the mere logical possibility of conflict (most of our concepts could be regarded as confused on this impossibly high standard); actual conflict would suffice, though it would have to occur in a number of cases sufficient to eliminate mere technical failures of apparatus and inexplicable breakdowns. Secondly, if the EEG criterion is to be confirmed as valid, it must be shown that there is some intimate causal connection between cessation of brain waves and the normal criteria; such a causal connection is particularly obvious in the case of criterion (c). This would establish that the disappearance of brain activity bore a more significant relation to death than mere interesting correlation, and would refute anyone who might allege that a blank EEG merely had been adopted as a convenient *convention* for determining death (*cf.* Malcolm, *op. cit.*, p. 76). Similarly, the fact that a physiological criterion of pain or of dreaming did not conflict with the ordinary criteria would be highly important in establishing it as a good criterion.

---

[1] Richard Rorty, 'Mind-Body Identity, Privacy, and Categories', *The Review of Metaphysics*, Vol. 19, 1965–66, pp. 24–54; reprinted in Stuart Hampshire (ed.) *Philosophy of Mind* (Harper and Row, 1966). In any case, Baier's attempt to make the pain case a special one is faulty, as I have shown incidentally in a recent article in *Mind*.

It might be said here (*pace* Malcolm) that the introduction of a different criterion would alter the concept of death, pain or dreaming beyond recognition. The concept would be altered, no doubt, but the overlap of the applicability of the new criterion with the applicability of the old would show the introduction to be an inoffensive move with a justified use of the same word. There are many instances of this kind of move in the history of science, where the same word is used for a scientific concept which differs from the ordinary one; for example, heat, curvature and force.

So much for the contention that the argument proves too much. Let us now see why it is ill-founded. Baier and Malcolm seem to be committing two errors. The first is to wield uncritically the notion of logical possibility, a notion which has caused so much trouble in philosophy that one expects it to be used now with more caution. The second is to operate with a rigid criterion/symptom distinction which admits of no fluidity in actual examples. Since this distinction stems from Wittgenstein (whom Malcolm acknowledges as his source), one has a right to expect *his* conception of the terms involved in it to dictate their use:

> In practice, if you were asked which phenomenon is the defining criterion and which is a symptom, you would in most cases be unable to answer except by making an arbitrary decision *ad hoc*. . . . Doctors will use names of diseases without ever deciding which phenomena are to be taken as criteria and which as symptoms; and this need not be a deplorable lack of clarity. . . .
> This [*i.e.* thinking in terms of the antithesis between criterion and symptom] is a very one-sided way of looking at language. . . . We are unable clearly to circumscribe the concepts we use; not because we don't know their real definition, but because there is no real 'definition' to them (*Blue Book*, p. 25).

> The fluctuation of scientific definitions: what today counts as an observed concomitant of a phenomenon will to-morrow be used to define it (*Philosophical Investigation*, I §79).

University of Western Australia

THREE MISTAKES ABOUT RETRIBUTIVISM

By Jeffrie G. Murphy

RETRIBUTIVE theories of punishment maintain that criminal guilt merits or deserves punishment, regardless of considerations of social utility. Such theories may be put forth for either of two reasons: (1) It could be argued (*e.g.* by a Moral Sense theorist) that the claim is a primitive and unanalysed proposition which is morally ultimate. Every ethical theory necessarily involves at least one such primitive (*e.g.* 'happiness is good', 'freedom is to be respected', *etc.*) and the retributivist may be offering this as his candidate. We can, he may argue, just *intuit* the "fittingness" of guilt and punishment. (2) It might be maintained (as it was, I believe, by both Kant and Hegel) that the retributivist claim is demanded by a general theory of political obligation which is more plausible than any alternative theory. Such a theory will typically provide a technical *analysis* of such notions as crime and punishment and will thus not regard the retributivist claim as an indisputable primitive. It will be argued for as a kind of theorem within the system.

The objection to the first sort of theory is obvious: the retributivist may be able to intuit the fittingness of guilt and punishment, but most of us cannot—not, at any rate, in a sense strong enough to make us want to appeal to the notion in justifying punishment. Thus the first theory is subject to all the classical objections to intuitionism, and these do not have to be repeated here. Let us, then, try to make sense of a theory of the second sort. What sort of theory of political obligation would render it plausible?

Consider a quasi-contractual model (found in Kant and John Rawls)[1] that seeks to analyse political obligation in terms of *reciprocity*. Such a model will proceed as follows: In order to enjoy the benefits that a legal system makes possible, each man must be prepared to make an important sacrifice—namely, the sacrifice of obeying the law even when he does not desire to do so. Each man calls on others to do this, and it is only just or fair that he bear a comparable burden when his turn comes. Now if the system is to remain just, it is important to guarantee that those who disobey will not thereby gain an unfair advantage over those who obey voluntarily. Criminal punishment thus attempts to maintain the proper balance between benefit and obedience by insuring that there is no profit in criminal wrongdoing. This is at least one point behind the

---

[1] See, for example, Rawls' 'Legal Obligation and the Duty of Fair Play' in *Law and Philosophy*, edited by Sidney Hook, New York, 1964. For Kant's views see his *Metaphysical Elements of Justice* (translated by John Ladd, Indianapolis, 1965) and my *Kant: The Philosophy of Right* (London, 1970).

jus talionis (return like for like) principle and was no doubt at least part of what Kant had in mind when he spoke, in a misleading idiom, of the criminal as owing a *debt* to the law-abiding members of his community. (The idiom of owing and paying a debt is misleading in that it tends to obscure the fact that (i) criminal "debts" differ from ordinary debts in that we have an antecedent moral obligation not to incur them and (ii) undergoing punishment for (say) murder, unlike paying the final instalment on a loan, can hardly be said to make things all right again, to make the world morally the same as it was before.)

Now the above theory is one possible form retributivism might take and is, at least, not absurd on its face. What I want to do now is to show how retributivism, at least of this variety, does not fall to three stock objections which are frequently put forth to "demonstrate" that no reasonable man could be a retributivist.

1. Retributivism as necessarily involving utilitarianism and as being obviously unacceptable without it

Rule utilitarians often maintain that retributivism, to be coherent, must involve a tacit appeal to utility—*i.e.* must tacitly presuppose that the principle 'Do not allow men to profit from their criminal wrongdoing' has more desirable social consequences than any alternative principle. But I should argue that Kant's theory, for example, is (i) perfectly coherent and (ii) quite independent of utilitarian considerations. His principle is that no man should profit from his own wrongdoing, and retribution attempts to keep this from happening. If a man does profit from his own wrongdoing, from his disobedience, this is *unfair* or *unjust*, not merely to his victim, but to all those who have been obedient. Now it may be, as the utilitarian might argue, that such unfairness—if widespread—would have socially undesirable consequences. But this is not Kant's argument. His argument is that the *injustice* or *unfairness itself*, regardless of consequences, demands retribution. As H. L. A. Hart has argued, 'a theory of punishment which disregarded these moral convictions [about justice] or viewed them simply as factors, frustration of which made for socially undesirable excitement is a different kind of theory from one which *out of deference to those convictions themselves* [justifies] punishment . . . ' ('Murder and the Principles of Punishment', in *Punishment and Responsibility*, Oxford, 1968, p. 79). Kant's theory is clearly of the latter sort.

But if this line is taken, the utilitarian may argue, retributivism becomes obviously unreasonable—a bit of primitive, unenlightened and barbaric emotionalism. But why is retributivism so condemned? Typically the charge is that infliction of punishment, with no attention to utility, is pointless vengeance. But what is meant by the claim that the activity in question is pointless? If 'pointless' is to be analysed as 'disutilitarian', then the whole question is being begged. One cannot refute a retributive theory merely by noting that it is a retributive theory and not a utilitarian theory. This is to confuse refutation with redescription. That the maximization of social utility is important is no more *obviously* true than that a man should not unfairly profit from his own criminal wrongdoing; and, if the utilitarian proposes simply to dig in his heels on the former, it is important to note that the charge of emotionalism cuts both ways.

2. The inapplicability of jus talionis

Perhaps the most common criticism of Kant's theory is the claim that the principle *jus talionis* (return like for like) cannot with sense be taken literally. As Hegel observes, 'it is easy enough . . . to exhibit the retributive character of punishment as an absurdity (theft for theft, robbery for robbery, an eye for an eye, a tooth for a tooth—and then you can go on to suppose that the criminal has only one eye or no teeth)' (*Philosophy of Right*, translated by T. M. Knox, Oxford, 1952, p. 72).

But this objection, as Hegel rightly sees, is superficial. Surely the principle *jus talionis*, though requiring likeness of punishment, does not require *exact* likeness in all respects. There is no reason *in principle* (though there are practical difficulties) against trying to specify in a general way what the costs in life and labour of certain kinds of crime might be, and how the costs of punishments might be calculated, so that retribution could be understood as preventing criminal profit.

3. The gap between theory and practice

Another common criticism against the Kantian theory may be regarded as Marxist in character. Kant's theory, it may be argued, involves an ideal utopian model of society which is in fact so utterly different from the actual character of society as to render it useless in understanding or evaluating any existing practice of criminal punishment. Indeed, the theory is dangerous. For it allows us to hide from ourselves the vicious character of actual social arrangements and thereby perpetuate gross injustice.

Let me elaborate: Punishment as retribution (paying a kind of "debt" to one's fellow-citizens) makes good sense with respect to a community of responsible individuals, of approximate equality, bound together by freely adopted and commonly accepted rules which benefit everyone. This is an ideal community, approximating what Kant would call a kingdom of ends. In such a community, punishment would be justly retributive in that it would flow as an accepted consequence of accepted rules which benefited everyone (including, as citizen, the criminal). But surely existing human societies are not *in fact* like this at all. Many people neither benefit nor participate but rather operate at a built-in economic or racial disadvantage which is in fact, if not in theory, permanent. The majority of criminals who are in fact punished are drawn from these classes, and they utterly fail to correspond to the model which underlies the retributive theory. Surely we delude ourselves in appealing to the retributive theory to justify their punishment.

The moral doubts raised here are extremely important. However, they may be doubts which testify to *strengths* rather than weaknesses in a retributive theory of the Kantian variety. Decent men surely want to object to the wanton handing out of punishments to those who, in a socially uneven community, always get the short end of the stick. But does not Kant's theory explain (or at least give one good reason) *why* we do want to object? Just punishment rests upon reciprocity; and is not one of the most serious moral problems confronting most existing communities the absence of such reciprocity, the absence of a balance between benefit and burden? Punishment is unjust in such a setting because it involves pretending (contrary to fact) that the conditions of justified punishment are met. Thus could not Kant, given his theory, easily share the Marxist scepticism about punishing in certain actual states? I believe that he could.

Now Kant may have thought that in defending a retributive theory he was also defending all of our present punitive practices. If so, his mistake is simply about *facts* and not about the *principles* involved. In developing his theory, Kant is attempting to outline the characteristics that a society must have if punishment within it is to be justified. Being right on this question will not prevent one from mistakenly believing that a particular society has these characteristics when in fact it does not.

My purpose here has not been to claim that a retributive theory of the Kantian variety is obviously true. I have rather been concerned to show that it is not, as is so often supposed, obviously silly. It is a *theory*, not just a primitive bit of intuitive vindictiveness, and should be taken seriously and criticized as a theory. It is not to be refuted by stock arguments like the above three; for these rest upon nothing but misunderstanding and question-begging.*

* Many of the arguments in this article have been taken from my paper 'Kant's Theory of Criminal Punishment', forthcoming in the *Proceedings of the Third International Kant Congress* (scheduled for publication in 1971). In the Congress paper, I was mainly concerned to pursue the topic of retributivism in a way that would be of interest to Kant scholars, and in the present article I have attempted to extract arguments of a more general philosophical interest. Also, in the Congress paper, I considered shared guilt and Christ's remark 'Let him who is without sin cast the first stone' as objections to retributivism. These objections, unlike the ones considered in the present article, I do not regard as obviously mistaken. For helpful discussion of the issues involved in both papers, I am grateful to Robert Gerstein, Gareth Matthews, Herbert Morris, and A. D. Woozley.

University of Arizona

ON DEATH AS A LIMIT

By JAMES VAN EVRA

PHENOMENALLY speaking, dying is simply a matter of ceasing to think and experience, and death presumably is the state of such experiencelessness. But while it seems perfectly reasonable to describe death in this way, to do so invites some potentially troublesome questions. For instance, it might be asked whether there need be anything which (phenomenal) death as a state characterizes. In describing death as a state, that is, are we tacitly committed to accepting the personal survival of death, in the sense that we need *something* of which we can say that it once thought and experienced, but, while still existing, does no longer? On the other hand, if we do not recognize the existence of experienceless selves, can sense be made of talking about phenomenal death as a state at all?

In this paper, I shall develop a view in which death is seen to be a state which characterizes absolutely nothing, and hence requires no commitment to the belief in selves which survive death. At the same time, the concept of death will be shown to retain all of its "ordinary" significance.

The heart of the view to be presented is an elaboration of Wittgenstein's view that death is like a limit, in the sense that we can approach it, but cannot reach it. The elaboration consists in pointing out that, like states which serve as hypothetical limiting constructs in science, death gains what significance it has, not by serving as a state characterizing things, but as a *function* which orders members of the series limited. In this way, the significance of the concept of (phenomenal) death can be retained, while the need to recognize the existence of things characterized by death can be rejected.

The paper is divided into three parts: first, Wittgenstein's basic analogy will be considered. This is followed by a discussion of the extension of the analogy, just outlined. Finally, the significance of the expanded view for more traditional views of death will briefly be considered.

I

In a terse comment entirely typical of the *Tractatus*, Wittgenstein says that 'Our life has no end in just the way in which our visual field has no limit' (6.4311). The analogy contained in Wittgenstein's remark should be understood as follows: Our visual field has no limit in the sense that there is no seeable bound to the visual field. Try as I may, I cannot see the limit, for to see something is to place it in the visual field, which would necessarily place it within the limits of the field. It could not, therefore, be the limit itself.

Of course, saying that, from the point of view of sight *simpliciter*, the visual field has no limit is not to say that the visual field is not limited in *any* sense. I infer, quite consistently with the fact that I cannot see the limit of the visual field, that my visual field is limited in a different sense (hereinafter called 'limited*') by, for instance, making inferences from the *manner* in which things appear in my visual field, as well as from evidence obtained from the other senses. That is, as things pop into and out of my visual field with some measure of consistency and coherence, it becomes useful to account for such intermittence by holding that, somehow, these things persist outside of my immediate field of vision. And because I can add to this evidence of the existence of such things from the other senses, the firm conclusion that my visual field is limited* is easily drawn.

Pressing the analogy, then, if my life has no end in *just the way* that my visual field has no limit, then it must be in the sense that I can have no experience of death, conceived as the complete cessation of experience and thought. That is, if life is considered to be a series of experiences and thoughts, then it is impossible for me to experience death, for to experience something is to be alive, and hence to be inside the bound formed by death.

Holding that my life has no end in this sense is perfectly consistent with the belief *that* I will die, in the same way that holding that my visual field has no (seeable) limit is perfectly consistent with the belief *that* my visual field is limited*. For just as I have grounds other then sight for believing that my visual field is limited*, I have grounds other than thought and experience as such for the belief that I will permanently cease to think and experience. I might, for instance, arrive at such a belief by realizing that I experience, from time to time, bodies which have ceased to function physiologically. From this I infer, on the basis of the analogy which I draw to my own (physiological) self, that I too will cease to function physiologically. Then on the basis of the belief that certain physiological functions are necessary for the occurrence of thought and experience, I infer that I will cease to think and experience, *i.e.* that I will die in this particular sense.

In sum, then, Wittgenstein is saying that within the phenomenal frame of reference, death is unexperienceable as the limit of the visual field is unseeable.

II

The elaboration of the analogy can be arrived at by first considering the following question: does the fact that I make use of the concept of the limit of the visual field depend on the existence of the limit? Need there be, that is, a field which is identical with my visual field in every respect except that it is inherently impervious to vision? To deny the reality of such a limit is, of course, not to deny that things or states of affairs exist unseen, for, as previously noted, independent grounds can be found for admitting their existence. Rather, to deny the reality of the limit is just to deny that there is a non-visual "field" which is complementary to my visual field but necessarily resists sight. Furthermore, were the limit, *i.e.* the complementary field, proved to exist (in some sense not involving vision) or if it were shown not to exist, the total effect on my vision would be nil, in the sense that nothing at all need change, regardless of the ontic status of the complementary field.

Since the supposed reality (or lack of it) of the limit of the visual field makes no (visible) difference, this suggests that questions concerning the visual reality of the limit are wrongheaded. In fact, to ask whether the limit itself is real *is* basically to misunderstand the significance of the concept of a limit. The answer to the question posed in the preceding paragraph is that to assert that the visual field has a limit need not be construed as a claim for the existence of an unseeable complement to the visual field. Statements about the limit of the visual field can far better be understood as stating relational properties between various "seeings" within the visual field. When I say, for instance, that I cannot see *X* as clearly as *Y* 'because *X* is closer to the limit of my visual field', such a statement employs the concept of the limit as a device with the use of which visual experiences are ordered or related. In such a case, reference to the limit need not be taken as an assertion of the existence of an (unseeable) state of affairs; the *total* significance of the statement can be accounted for solely in terms of the contents of the visual field and their properties, and no acceptance of the limit as something super-added is required. The significance of the limit is not as something independently real, but as an operational device.

Since talking about the visual field may be somewhat confusing, due to the fact that we are unavoidably close to the topic at issue, I should like to make the same point with regard to a more neutral subject matter —limits as they are used in science. As an example of such a limit, consider a modern reformulation of Newton's First Law of Motion[1]: 'If the external forces F acting on a body (whose momentum along a straight line is *mv*) are zero, then the time-rate of change in *mv* (which may happen to be zero in the limiting case, so that the body is at rest relative to that line) is also zero'.[2] Formulated in this manner, the axiom postulates a *limiting* motion, *i.e.* a state in which a body is under the influence of no forces at all. To ask whether there are *really* bodies in such a state is to miss the significance of the concept entirely. Nagel makes this point clear in the following passage:

---

[1] Which in its original version is as follows: Every body continues in its state of rest or of uniform motion in a right line unless it is compelled to change that state by forces impressed upon it.

[2] Nagel, Ernest, *The Structure of Science*, New York: Harcourt, Brace and World Inc. 1961, p. 159.

> The first axiom formulates a complex set of facts in terms of a postulated *limiting* motion, were the series ideally prolonged without limit. However, the axiom must not be read with a sort of myopic literalness; it should not be construed as asserting that there are in fact bodies under the action of no forces, or as requiring for its validity the existence of such bodies. For the language of limits must be handled with care. In physics, as in mathematics, the assertion that a series of terms has a limit is often best construed as simply a way of stating a relational property characterizing the unquestionably existing members of the series, rather than as a statement which affirms the (possibly doubtful) occurrence of some term not initially assumed to be a member of the series. Accordingly, the first axiom does have an empirical content, for the axiom formulates certain identifiable relational characteristics of the actual motions of bodies, all of which are under the action of forces, when these bodies are serially ordered (p. 184).

It might be added that, if there *were* a body under the influence of no forces, this would neither add to, nor detract from, the significance of the concept of limiting motion from the point of view of the series of bodies under the influence of forces. *All* of the significance of the concept as a limit derives from its use as an ordering function within the series of bodies under the influence of forces. If the limiting state is "real" it is simply not in a sense which is continuous with that of the reality of the series of bodies under force.

Furthermore, were it shown that no body *could* be under the influence of no forces, this would not mean that the series of bodies under the influence of force is not limited. I can believe, that is, that the series is limited without having to believe in the existence of a body under the influence of no force. This is even clearer in the case of another limit, absolute zero. The Third Law of Thermodynamics, if true, insures that no body can possibly have a temperature of zero degrees absolute. Yet this does not detract in the least from the significance of the limit as an ordering device for things-having-temperatures. Again, the series' being limited is completely compatible with the fact that nothing exists at absolute zero.

So far, only the "sight" side of Wittgenstein's analogy has been extended. Now the question arises, if the limit of the visual field is best treated (visually) as an ordering function, rather than as a state of affairs characterizing something, might not this treatment profitably be extended to the other side of the analogy, *i.e.* to death as a limit? I should like to answer in the affirmative, by suggesting that the concept of death as a state functions for us in precisely the same manner in which all of the previously mentioned limits function in relation to the members of the series which they limit. Death *qua* the cessation of thought and experience is not *just* a bound, although it is in itself conceptually unattainable, just as absolute zero is not *just* a limit although nothing can exist at such a temperature. In each case, the importance of the limit, and virtually *all* of its significance, derives from the fact that the limit serves as an ordering device.

To say that death functions as a limit in this sense is simply to say that we build, to a considerable extent, our conceptual lives with the *use* of it. The general, usually unconscious, realization that we are approaching the limit often impels us to, *e.g.* truncate projects so that we can go on to others; it forces us to pick and choose those tasks from the multitude which are more "worthwhile". In general, we form a "set" *i.e.* a context from which life is viewed *via* the realization that life is limited, and that we are approaching the limit. All of this can be accomplished without any conception of an experienceless, literally thoughtless "self" whatsoever.

While there are many specific examples of the "mechanics" of death considered as an analogue to a limit, I shall mention only two. First, death sometimes intrudes in our lives with particular vigour. When someone slows his driving from the awareness that his driving could eventuate in a hastily arrived at death, he employs the concept by using it as a device which assists in the ordering of his conceptual affairs. In such a case, the individual may of course be able to conjure an image of what being involved in a crash would do to him physically, but this is not death as a limit and it does not get him closer to being able to conceive a state of experiencelessness. He cannot conceive of his (thinking and experiencing) self in a state characterized by the complete inability to think and experience, yet he reacts to precisely the possibility of attaining such a state. This again is analogous to conceiving of the temperature scale as a "reaction" to absolute zero, while existing at absolute zero is impossible.

The second and purest example of how we react to death has to do with the reactions of those who are aware that their death is imminent. A recent psychiatric study[1] suggests that individuals react to impending death in fairly regular ways. In each of the stages through which such individuals pass (denial, anger, bargaining, depression and acceptance) the individual conceives of death, not in itself, but as it bears on him while he is still alive. Death to these individuals is something to be avoided, to blame others for, to reject others in the face of, while it steadfastly resists conceptualization itself.

Because we react to death in these and many other ways, does this mean that we are implicitly committed to the existence of a self which is like our present one but cannot think or experience? Do we need, that is, some continuing "something" of which we say that it is now alive (*i.e.* is experiencing) and later, while persisting, will not be experiencing? The answer is no; no more than using the concept of absolute zero commits one to the existence of a body at that temperature, or using the concept of bodies under force commits one to the existence of bodies under no force. The significance of the concept of death as a limit is completely exhausted in its function as an ordering device; it requires no assumption of the existence of some super-added experienceless self.

---

[1] Kübler-Ross, Elizabeth, M.D., *On Death and Dying*, NY: The Macmillan Co., 1969.

In sum then, we have seen that, as Wittgenstein suggested, death is like a limit in the sense of being an experientially unattainable bound. But more than this, death is like a limit in deriving what significance it has from the role it plays as an ordering function with respect to the experiencing beings which it "limits".

III

Finally, I should like briefly to point out how the view of death here presented bears on other views of death.

First, the present view is strongly opposed to the view, held by many continental philosophers from Schopenhauer to the present, that we have an "awareness" of death *as such*, and that from this we come to regard death as an evil, the recognition of which points up the tragedy of the human condition, making total human happiness impossible to achieve. The view I have advanced exonerates death as the purported snake in our garden. We have, as experiencing beings, no more an "awareness" of death as such than we have, as beings which have temperatures, an awareness of absolute zero. Limits like death (*i.e.* limits on our powers of conception) are simply out of our league—we cannot "grasp" such limits. They are contentless and hence provide absolutely nothing for our awareness. From this point of view, death cannot diminish our human happiness, in the same way that bodies-in-motion are none the worse for having a lower bound. The bound, not being a member of the series, cannot defile it. The series is what it is, happy or unhappy, good or bad, quite independently of any bound as such.

Secondly, the view presented provides an explication of Epicurus' remark that death is nothing, 'either to the living or to the dead, for with the living it is not and the dead exist no longer'. Or rather, it explains the fact that his remark seems strange. Epicurus is speaking of the limit as such, and he is surely correct in saying what he does about it, particularly because it is aimed at those who hold that death is to be feared in the sense of fearing what might happen to the "self" which goes on after death. The reason why his remark seems odd is that it purposely ignores the sense in which death can be said to be *something*—which is the sense which captures all that death as a limit is. Conceiving of death as a function of life actually aids Epicurus' criticism by showing how talk of death can be so construed that the assumption of the existence of experienceless selves can be avoided.

Finally, the present view of death aids in the avoidance of pitfalls inherent in what might be called the "everyday" view of death. Consider the following formulation of such a view, provided by Paul Edwards:

> It seems quite plain that human beings not infrequently imagine and conceive of their own deaths without the least difficulty, as, for example, when they take out life insurance or when they admonish themselves to drive more carefully. Nor is it at all difficult to explain what a person imagines when he thinks of his own death.
> 'When I die,' wrote Bertrand Russell in a famous passage (in 'What I believe'), 'I shall rot and nothing of my ego will survive'; and it is surely this that people wish to avoid or put off. A person thinking of his own death is thinking of the destruction or disintegration of his body and of the cessation of his experiences.[1]

While comfortable and ordinary, talking in this way about death is clumsy and can easily lead to quite undesirable consequences. We cannot imagine or conceive of our death as such if death is considered to be the cessation of experience. And while I might, as suggested earlier, imagine my dead *body* while driving recklessly, I imagine nothing of death *itself* when I take out life insurance. Rather, I am much more likely simply to imagine what would happen to those who depend on me were it the case that I could no longer provide for their welfare. Furthermore, it is not the fact of the cessation of the ego as such that we wish to avoid or put off, because this is simply out of our conceptual range. We cannot even *wish* to avoid it if we cannot conceive it in itself. What we wish to avoid, rather, are all of the results of applying death as a functor: separation from family and friends, inability to attain goals sought, inconvenience to others, *etc.* A person thinking of his own death simply cannot think of the cessation of his experience, unless this is an ellipsis for saying that he is thinking of the results of using *death* as a limit.

In summary, death treated analogously to a limit literally *is* the manner in which it is used, and to treat it in any other way is to miss its significance.

University of Waterloo

---

[1] Paul Edwards (ed.) *Encyclopedia of Philosophy*, New York, The Macmillan Co. and The Free Press, 1967. Vol. 5, p. 416.

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, Miss M. A. Boden; J. H. Benson, Alec Fisher, A. Harrison, R. F. Holland, S. Körner, A. C. MacIntyre, A. M. MacIver, B. Mayo, R. G. Swinburne, A. R. White. This committee is responsible for appointing and advising the editor and for the general policy of the paper.

19 DEC 1971

SUBSCRIPTIONS. The subscription to ANALYSIS, Vol. 31, is £1.20 in Great Britain and £1.40 abroad (U.S.A. $4.00), post free. Each volume comprises six numbers, five of 32 pages and one of 48 pages, appearing within the academic year—in October, December, January, March, April and June. Orders should be sent to Basil Blackwell, 49 Broad Street, Oxford, or placed with any bookseller.

CONTRIBUTIONS. Articles submitted for publication should be addressed to **Dr. C. J. F. Williams, 1 Fossefield Road, Midsomer Norton, Bath BA3 4AS.** Contributors are asked to note the following requirements.

Articles should **normally** not exceed 3,000 words in length. **Occasionally,** however, longer contributions can be accepted;

They must be typewritten in double spacing on one side of the paper only;

Footnotes should be kept to a minimum and wherever possible avoided altogether;

Single quotation marks should normally be used, except for purposes of internal quotation and "scare" quotes.

Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

Contributors in the United Kingdom should enclose a stamped addressed envelope of suitable size; if immediate acknowledgment is required a stamped postcard should also be enclosed.

Overseas contributors who wish to have their MSS. returned should send an envelope **and international reply coupons** of the requisite value, whether for air or surface mail.

Proofs are normally read by the Editor. A contributor who wishes to read his own proofs should notify the Editor at the same time as the MS. is submitted.

OFFPRINTS. Contributors will be supplied free of charge with a limited number of offprints of their articles. Further copies may be ordered at a charge from the publishers. Orders for these, if required, should be placed **direct with the publishers** when an article has been accepted for publication.

The copyright of articles printed in ANALYSIS remains the property of the author, but contributors are strongly advised, in their own interest, to consult the Editor before consenting to the reprinting of their articles.

PRINTED IN GREAT BRITAIN BY A. T. BROOME AND SON, 18 ST. CLEMENT'S, OXFORD

ol. 31 No. 6 (New Series No. 144) June 197

ANALYSIS

Edited by
PETER WINCH

CALCUTTA UNIVERSITY
CENTRAL LIBRARY

CONTENTS

On behalf of the Skolemite — WILLIAM J. THOMAS

The expressive capacity of non-translational languages — HANS G. HERZBERGER

Popper's verisimilitude — G. S. ROBINSON

The infallibility of our knowledge of our own beliefs — U. T. PLACE

On not being said to do two things — GARETH B. MATTHEWS

BASIL BLACKWELL · BROAD STREET · OXFORD

Twentyfive pence

ANALYSIS 31.6 JUNE 1971

CALCU
CENTR
LIBRA
UNIVE

ON BEHALF OF THE SKOLEMITE*

By WILLIAM J. THOMAS

MICHAEL David Resnick has raised some interesting new objections to the position he calls "Skolemism".[1] "Skolemism" is to be understood, roughly, as that position about mathematical objects which denies the existence of absolutely uncountable sets (Resnick, 1969, p. 185). That is to say, a "Skolemite" is one who maintains that, on grounds provided by the Löwenheim-Skolem Theorem, no set fails to be countable in some formal system.

In an earlier attack on Resnick's view that there are absolutely uncountable sets, I argued that the assumption that there are no such things is a perfectly consistent view, and that the Löwenheim-Skolem Theorem shows that it is consistent (ANALYSIS, 28.6,1968). It was also argued there that there are good reasons to prefer the Skolemistic position to Resnick's Platonistic one. (I suggested that the use of ordinary language as a universal metalanguage is the essential characteristic of Platonism in mathematics. It is in this sense that I use the term here.) Resnick has conceded the former point (*ibid.*, p. 195), but he maintains in his latest paper on the subject that there are still good reasons to embrace Platonism in the matter of uncountable sets. In the present paper Resnick's new arguments will be examined and another brief apology for Skolemism will be offered.

It may be useful to review some of the key notions and results pertinent to any discussion of the philosophical significance of the Löwenheim-Skolem Theorem. We shall use the word 'theory' to mean just an effectively specifiable set of sentences. We understand an interpretation of a theory to be a specification of meanings for the terms of the theory, terms being just the referring expressions (if any), and the predicates and functors (if any). We speak of a theory having a model just in case the theory can be interpreted in such a way as to be true of the objects (which collectively are called the 'domain' of the interpretation or model) which are the *referentia* of the interpretation.[2] A theory is consistent if and only if it has a model.

Intuitively, any set, and a domain in particular, is countable if and only if there is a natural number for each thing in the set. A little less intuitively, we say that a set S is countable just in case there exists a one-to-one function from S to the natural numbers. We say that a set is uncountable if and only if there is no such function.

---

* This work was supported by a Summer Research Grant from the Foundation of the University of North Carolina at Charlotte.

[1] Resnick's first discussion of this subject is 'On Skolem's Paradox', *Journal of Philosophy*, LXIII, 15, (1966). His second and latest paper on the subject is 'More on Skolem's Paradox' *Noûs*, III, 2, (1969).

[2] A full account of these matters is available in: Elliott Mendelson, *Introduction to Mathematical Logic*, (New York: Van Nostrand, 1964), pp. 49–56.

More formally, we say that a set S is countable in a theory T if and only if it is a theorem of T that there is a function of the kind described above, for S. We say that S is weakly uncountable in T if and only if S is not countable in T; and finally, S is uncountable in T if and only if it is a theorem of T that there is no function of the kind required for S to be countable. Such a function shall hereinafter be referred to as an 'enumerating function'. The Löwenheim-Skolem Theorem says that if a theory T has a model then it has a model with a countable domain.

It should be noticed that the same set[1] may be countable in one theory and uncountable in another. The Löwenheim-Skolem Theorem provides examples of just this phenomenon, since it says, essentially, that if there is a theory T′ in which a given theory T is interpreted so as to provide a model for T, then there exists another theory T″ which also provides a model for T and which does so with a domain which is countable in T″. This domain will not in general be countable in T, however. For theories T in which Cantor's Theorem (which states that there are uncountable sets) is provable, the model provided by the Löwenheim-Skolem Theorem cannot be countable in T, for in that case T would be inconsistent, and would, therefore, have no model.

The Skolem Paradox arises when we fail to make a rigorous distinction between object and metalanguages while at the same time allowing both Cantor's Theorem and the Löwenheim-Skolem Theorem. In particular, the danger of the Skolem Paradox becomes especially great when we talk as though we had available one universal metalanguage in which all theories could be discussed. This situation sometimes arises from too free a use of ordinary language in discussions of metamathematics.

We are now in a position to undertake discussion of Resnick's latest objections to Skolemism. We will not concern ourselves here with any of his earlier objections, nor with those of other philosophers.

The first of Resnick's new objections is an objection against a certain argument for Skolemism, rather than an objection against the position itself. It has been argued that models, like the formal systems of which they are the "subject matter", must be formally specified.[2] One basis of this argument is that paradoxes, the Skolem paradox in particular, arise when we permit ourselves to luxuriate in the sort of naive talk of sets which is required to obtain 'uncountable' informally specified models. Another basis of this argument is that if the anti-Skolemite is to argue that Skolemism is wrong because it falls to map "the real numbers" into the natural numbers (*cf.* Resnick, 1966). It is just not cricket to assume the existence of an (informally specified) set of all real numbers, since the very existence of such uncountable sets is precisely the matter in question.

---

[1] As is argued in Resnick, (1966), there are problems about just how to construe claims of intersystematic set identity. However, as is argued in Thomas, (1968), Wang's system Σ provides an adequate reconstruction of intersystematic set identity in which the claim I make here can be verified.

[2] The argument may be found both in: Arthur Fine, 'Quantification Over the Real Numbers', *Philosophical Studies*, XIX, 1–2, (1968), and Thomas, (1968).

Resnick's reply to the Skolemite's rejection of informally specified models is the charge that without informally specified models the Skolemite cannot sustain his countability claims about formally specified sets. Resnick tells us that '. . . the Skolemite is in trouble with even the natural numbers—his one sure source of countable domains for his countable models' (Resnick, 1969, p. 187). The cause of the trouble is, on Resnick's account, the failure of *categoricity* of (first order) formal theories of the natural numbers. Intuitively, what this means is that any (first order) theory of the natural numbers will have (either no model, or) at least two models which differ so greatly in logical structure that it would be unreasonable to call both of their domains "the natural numbers". Thus, Resnick maintains, one cannot claim to have established the countability of some set in an entirely formal manner. For to establish countability of a set is to establish the existence of a certain function from that set to the natural numbers, and, on a purely formal basis, Resnick argues, one cannot identify the natural numbers.

I think one might reply to this criticism of Skolemism's rejection of informal models by denying Resnick's tacit assumption that the theories in which models are specified must be *first order* theories. It is only to first order theories of the natural numbers that the failure of categoricity theorem (to which Resnick appeals) may be applied.[1] Indeed, if one allows oneself the apparatus of quantification over function variables, one can readily formally state Peano's Postulates, and it is known that Peano's Postulates *are* categorical.[2]

Adopting a more rigorous standard of ability to identify "the natural numbers", Resnick points to the phenomenon of "numerical insegragativity",[3] arguing as follows: Even Peano's Postulates suffer from numerical insegragativity, in the sense that if $N=\{0, 1, 2, \ldots\}$ is the domain of a model for the Postulates, then so is $N'=\{\alpha, 0, 1, 2, \ldots\}$, where α is anything whatsoever. (Here, as earlier, we adopt the convenient, standard, though not strictly correct, way of speaking about natural numbers: "*the* natural numbers". Cf: Quine 'Ontological Relativity', reprinted in *Ontological Relativity and Other Essays*, (New York: Columbia University Press, 1969), pp. 44–45. We also employ the familiar but potentially similarly misleading "listing" form of set description.) Then, Resnick argues, it suffices to notice that $N \neq N'$ to become convinced that one cannot, by formal methods alone, identify "the natural numbers".

---

[1] Resnick, (1969), p. 187. *Cf.* Mendelson, (1964), p. 116, Exercise 1.

[2] Mendelson, (1964), p. 116, Exercise 3.

[3] The term is from Quine, 'On ω-Inconsistency and a So-called Axiom of Infinity', reprinted in Quine, *Selected Logic Papers*, (New York: Random House, 1966), pp. 118–120.

There seem to me to be two possible responses that the Skolemite might make to this argument: He might provide some other formal specification of what is meant by Skolemistic countability claims, or he might, with Quine (*Ontological Relativity*, p. 45), take the line that "the natural numbers" can be just any set which satisfies the laws of arithmetic, and thus, either N or N′ can be taken to be the natural numbers. Since the second, Quinean, response is so well elucidated in several of Quine's writings, it seems unnecessary to discuss it further here. We turn then to the first sort of response.

Resnick is certainly correct that it would be at least a little perverse of the Skolemite to insist that there are no uncountable sets while having to maintain that there is a set of natural numbers. As a formalist, the Skolemite would have a difficult time in justifying claims of the existence of any particular set. Quine has provided the Skolemite with one escape from having to countenance the existence of particular sets, but it would be better to escape altogether from commitment to a set of natural numbers. The Skolemite may explicate what he means by countability in a purely formal way. I think we may take it that any theory of sets of any use either to the Skolemite or to the Platonist will contain the central core of Cantor's intuitions about sets, and, of course, that such a theory will be consistent. Then, we may define countability for a theory T as follows:

> Def. 1: A set α is Dedekind infinite in T *iff* it is a theorem of T that there is a proper subset y of α and α can be mapped one-to-one into y.
>
> Def. 2: A set α is countable in T *iff* it is a theorem of T that every Dedekind infinite set y is such that α can be mapped one-to-one into y.

Def. 2 gives a formalistic definition of countability[1] in a theory. I think it is clear that both of these definitions can be fully formalized in any adequate (in the sense mentioned above) set theory. I think it is also clear that they do the required task without having to presuppose any "set of all natural numbers".

The Skolemite is not any more committed (at least he ought not to be) to a transcendent set of natural numbers than he is to transcendent

---

[1] 'Countable' is sometimes applied only to sets which are countably infinite. Def. 2 defines countability as countably infinite or finite. This is the notion which the Skolemite needs, since he would be content to have finite models, as well as countably infinite ones.

Another approach to a formalistic definition of countability is to take a set to be countable in a theory if and only if it is provable in that theory that the set is not the power set of any infinite set. This approach, as was pointed out to me by Michael Jubien, requires that we assume the Generalized Continuum Hypothesis.

(or "absolutely") uncountable sets. If he were, then if he were consistent, he would have to deny the truth of the principle of comprehension of power set, which principle seems as good a candidate as any set theoretic principle to be called 'true'. The strength of the Skolemite's position seems to be his view that it is only correct to say of sets that they are uncountable or that they are countable in the context of some theory.

For the strict formalist an "absolutely uncountable set" would be, I suppose, a set which was uncountable in every consistent system. The Löwenheim-Skolem result does clearly show that there aren't any sets which are "absolutely uncountable" in this sense. Resnick's Platonistic view enables him to define an "absolutely uncountable set" as one which is uncountable in every correct (or true) theory (Resnick, (1969), pp. 189–192, especially, p. 191). The Skolemite would not quarrel with the claim that some consistent theories have models which are uncountable *in certain meta-theories*.[1] Further, there does not seem to be anything wrong with Resnick's use of such meta-theories to describe a model he views as "the intended model" of a given theory. Just as the Löwenheim-Skolem theorem shows the consistency of the view that there are no absolutely uncountable sets, so the "Hidden Inflation Theorem" shows the consistency of the claim that particular (consistent) theories have models which are uncountable (relative to some theories). The real issue would seem to be whether or not there is some reason to prefer the view that there are (Platonistically) absolutely uncountable sets, to the view that there aren't any such things.

Resnick's reasons for preferring the view that there are uncountable sets, in his absolute sense, fall into two general classes, which may be associated with the last two sections of his paper. The first class of reasons have to do with Resnick's views on the role of formal systems, and the second class with more or less technical problems he sees with the Skolemite's position. We turn now to the first, more philosophical, class of reasons.

> My view, which I share with its originator, Frege, is that formal systems are primarily tools for stabilizing, codifying and exploring pre-existing domains of mathematics and not for creating new ones. As a result, the intended model is already at hand when the formalism is introduced (Resnick, (1969), p. 189).

The formal system Resnick chooses as an example here is that of von Neumann, Bernays and Gödel, which, following Resnick, we will refer to as 'NBG_f'.[2] The idea, then, in the above quotation, is that there is a certain intended interpretation of NBG_f, the domain of which is uncountable, and which is informally specified. We are justified, on Resnick's account, in informally specifying that domain because we already have that domain "at hand" when first we begin the construction of NBG_f. Thus, on Resnick's account, the situation in set theory is much like that in a formalization of, say, Newtonian mechanics. We know what we expect of such a formalization in the first place, and so it is silly to ask whether the references of the completed formal theory are real or not.

---

[1] Actually, every consistent theory has models which are uncountable in some consistent theories. This is the force of the "Hidden Inflation Theorem". Cf: Quine, 'Implicit Definition Sustained', reprinted in *The Ways of Paradox and Other Essays*, (New York: Random House, 1966), pp. 196–197.

[2] Resnick, (1969), p. 189. 'NBG_f' is understood to be the (uninterpreted) formal system, 'NBG_u' its intended interpretation, and 'NBG_c' its countable Skolem model.

The issues of whether set theory is "true", whether there really are sets, and what the role of formal systems is, are such complex ones that it would be pretentious to attempt, in this brief paper, to attack Resnick's metaphysical presuppositions. He has given reasons for the Fregean view which he holds in earlier papers, and we are concerned here only with his latest attack on Skolemism. However, it does seem possible to offer some objections to his present arguments from those presuppositions.

There seems to me to be a serious problem in saying just what the "intended model" for set theory is. In his earlier paper on Skolemism, Resnick wrote as though he thought the domain of that model might be just the real numbers (Resnick, (1966), p. 431). Certainly any consistent set theory will have a model in the real numbers, but it seems far more plausible to say that if there is an *intended* model for set theory, it has as its domain V, the set of all sets. This indeed seems to be Resnick's present view (Resnick, (1969), pp. 190–191). The example of NBG_f chosen by Resnick is a particularly unfortunate one for his purposes, since it is a theorem of NBG_f that V does not exist![1] Thus if NBG_f is adequate to our informal ideas about sets, Resnick is wrong in thinking that the intended model exists. I suppose that this argument of mine may appear to be just a quarrel with Resnick's example; he could have chosen Quine's NF, in which V exists (*ibid.*, p. 277). However, even granting the availability of more suitable examples, the point may still be made that if the subject of set theory is *sets*, then it appears that set theories themselves provide the only clue as to what, beyond the most rudimentary properties, that subject matter is like. I would argue that we don't understand 'uncountable set' outside the context of some set theory, because that notion is an inherently set theoretic one. If one adopts a Platonistic informal meta-theory, one would seem to have taken on the full burden of Cantor's "paradise".

Resnick's view is that we don't create new mathematical objects when we devise new formal systems. But how can we be sure that the mathematical objects which set theory is intended to formalize are real? We might, for example, devise a formal theory of the genetics of centaurs. We could then, correctly it seems to me, say that the intended model of this theory has as its domain a certain set of genes. But, I think we would agree, there isn't really any such set, because there aren't really any centaurs, not even centaurs *qua* mathematical objects.

---

[1] We take '$V \in V$' to mean 'V exists'. Cf: Quine, *Set Theory and Its Logic*, (Cambridge, Massachusetts: Belknap Press, 1963), p. 45. Interestingly, $V \in V$ in Zermelo-Fraenkel set theory as well.

It appears that it is only when a certain domain is required by a given theory, *i.e.* when only a certain domain will serve to model a given theory, that we are willing to say that the acceptance of the theory commits us to accept that domain. When we accept, say, a domain of centaurs, it is for reasons other than model theoretic considerations. In mathematics, however, model theoretic considerations seem to be about all we have available on which to base our choices of ontologies. Richard E. Grandy, in 'On What There Need Not Be', *The Journal of Philosophy*, LXVI, 22, (1969), p. 812, suggests that one way to avoid wholesale Pythagoreanism is to adopt a less model theoretic view of ontology than has been implicit in most of the recent discussion of the ontological significance of the Löwenheim-Skolem Theorem. My remarks here are in the spirit of Grandy's suggestion, at least as that suggestion applies to extra-mathematical subject matter.

The difficulties with Skolemism, characterized above as "technical", which Resnick raises, have to do with the Skolemite's being required by his position to: (1) utilize strong formal theories, and (2) assume the existence of rich ontologies. Because of these requirements, Resnick argues, Skolemism, though consistent, offers no attractions over Platonism.

In regard to (1), Resnick tells us that

> . . . a detailed examination will show that (the case against uncountable sets) is as much an attack against the acceptance of strong formal systems such as NBG_f without regard to any associated model or ontology. This is quite clear with respect to the consistency problem. If NBG_f is inconsistent then neither my NBG_u nor the Skolemite's NBG_c exist.[1]

If I understand Resnick correctly here, he is arguing that if the rich ontologies which are the intended interpretations of strong formal systems (such as NBG_f) are problematical, then so are those formal systems used to describe those ontologies. I think, though, that the Skolemite can deny this conclusion by simply denying that there are those rich problematical ontologies. Why does so doing require him to "accept" NBG_f, or some even stronger theory?

Resnick answers this question by arguing that there are two general ways to obtain countable ("Skolem") models for NBG_f (or for any other theory to which the Skolem theorem is applicable), and that the formal meta-theories in which either of these ways may be realized are, in certain senses, stronger than NBG_f. The two ways are, basically, just the applications (to the set theory under consideration, be it NBG_f or some other theory) of the two principle versions of the Skolem theorem, one which requires the axiom of choice, and one which does not. In the latter form, the theorem gives us a countable model (provided, of course, that the theory under consideration is consistent), but that model has the feature that the relation R in the model which is the interpretation of the membership predicate 'ϵ' is neither the intuitive membership relation, nor is it a "familiar" number theoretic relation (*ibid.*, p. 194). I don't see why the Skolemite ought to be particularly upset by the interpretation of 'ϵ' since it is his contention that the "intended model" for set theory is unclear, and thus it ought not to be surprising that the sole primitive predicate of set theory should be modelled by an unfamiliar relation between natural numbers.

---

[1] Resnick, (1969), p. 194. The case against uncountable sets to which Resnick refers is well described by him on p. 193. It is the standard sort of argument from the paradoxes.

Resnick argues that the second form of the Skolem Theorem (without need of the axiom of choice) '. . . commit(s) the Skolemite to a highly artificial extension, of number theory which is formally stronger than NBG_f . . .' His argument is that

> . . . the only known axiomatization of (the metatheory in which the Skolemite's application of the Skolem Theorem, *sans* the Axiom of Choice, to NBG_f is made) involves the addition of 'CON(NGB_f)', the numerical consistency sentence for NBG_f, to the axioms of (a standard system of formal number theory) (*ibid.* p. 194).

There are several remarks to be made about this argument: The argument is in the form of an *ignoratio*, though perhaps that form of argument is justified here, since what Resnick really wants to establish is that there are no good reasons for holding the Skolemite position (*ibid.*, p. 195), and not quite the stronger claim that the Skolemite is committed to formal systems stronger than conventional set theories. Further, the comparison of strengths of formal systems, and set theories in particular, is not a matter of clear agreement among workers in foundations of mathematics. Moreover, if NBG_f is known to be consistent, then the axiom 'CON(NBG_f)', being known to be true, seems harmless enough. But if we lack good reasons for believing in the consistency of NBG_f then we lack good reasons for believing in NBG_u.[1] Hence, if Resnick's Platonistic doctrine is justified, then so is the acceptance of this axiom to which Resnick objects.

The application of the Skolem Theorem which requires the axiom of choice proceeds by using the axiom of choice on the domain of any model for a given theory to yield a countable subset of that domain. The countable subset of the original domain also models the system under consideration. This method has the nice feature that 'E' is interpreted in the countable domain by a relation which is the membership relation in that domain. Resnick objects to this Skolemistic application of the Skolem Theorem by saying:

---

[1] I think Resnick would concede this point. Cf: Resnick, (1969), p. 194.

> Unlike the first choice open to the Skolemite this method would commit him to both (1) a metalanguage rich enough to permit the discussion of NBG_f and all its models, and (2) the existence of some (possibly uncountable) model for NBG_f. The first commitment is necessary because in order to establish a difference between his model and NBG_u the Skolemite must specify D (the domain of his countable model). In other words, his metalanguage must be rich enough to carry out Skolem's proof. The second commitment is required because the *definiens* of D refers to the domain of some unspecified prior model of NBG_f. Thus once again the Skolemite is committed to a theory which is formally stronger than NBG_f (*ibid.*, p. 195).

The Skolemite can respond to the claim that he is committed to (1) by admitting that it is quite literally true. Certainly anyone who discusses in a serious way the consistency of a theory T is committed to a theory which is, in a precise sense,[1] stronger than T. If the Skolemite denies, as he in fact does, the existence of uncountable sets, then he denies the existence of uncountable models, and so by his lights, there would be no problem in countenancing a metal-theory for NBG_f in which one could refer to *all* of the models for that theory. The Skolemite's position is that those models would be countable. Further, inasmuch as theories are just effectively enumerable sets of rather well understood objects, *viz.* sentences, there does not seem to be any particular reason for the Skolemite to be embarrassed by his commitment to *any* theory, provided of course, that that theory is consistent. Resnick has conceded the consistency question (*ibid.*, p. 195). It is really the crux of the Skolemite's position that, in general, one's theories need have little to do with one's ontology.

Resnick's argument in regard to (2) is that if the Skolemite's countable model is defined in terms of some (unspecified, possibly uncountable) other model, then by being committed to having a model for NBG_f the Skolemite is committed to 'the existence of some (possibly uncountable) model for NBG_f'. But this problem is just a special case of a more general problem for all cases of "ontological reduction". That problem is that whenever we want to say that we don't need all of a certain universe, we have to speak about the superfluous entities. A fellow anti-Skolemite of Resnick's has provided an excellent answer to the general problem:

---

[1] This sense is given by "Gödel's Second Theorem" on the unprovability of CON(T) in T, for a wide range of T. Further light has been shed on the precise sense in which the consistency of T is unprovable in T in certain work of Feferman. The matter is well treated in a forthcoming paper by Resnick.

> This sounds . . . as if no ontological economy is justifiable unless it is a false economy and the repudiated objects really exist after all. But actually this is wrong; there is no more cause for worry here than there is in *reductio ad absurdum*, where we assume a falsehood that we are out to disprove. If what we want to show is that the universe U is excessive and that only a part exists, or need exist, then we are quite within our rights to assume all of U for the space of the argument. We show thereby that if all of U were needed then not all of U would be needed; and so our ontological reduction is sealed by *reductio ad absurdum* (Quine, (1969), p. 58).

The Skolemite does precisely as Quine suggests when he applies the second (*cum* Axiom of Choice) version of the Skolem Theorem to NBG_f: he shows that if all of NBG_n were needed then NBG_c would suffice.

I cannot claim to have established that Skolemism is true. In the present discussion I have tried to defend that position against Resnick's latest attack. If my defence has been successful, as I think it has, then the Skolemite's position has at least Ockham's dictum in its support. Also, I think that the reasons advanced in my earlier paper for holding the Skolemite position stand unrefuted.

University of North Carolina at Charlotte

THE EXPRESSIVE CAPACITY OF NON-TRANSLATIONAL LANGUAGES

By HANS G. HERZBERGER

TARSKI'S semantic conception of truth proceeds from the notion of satisfaction as a relation between the terminology and ontology of a language. Powerful as this conception may be, it has so far resisted implementation in connection with natural languages. Tarski himself has claimed this to be a matter of principle, so long as the classical logical structure obtains. Accordingly every language would have expressive gaps with respect to its own semantic theory; and no human language would be adequate to the formulation of a general theory of human language, barring some thoroughgoing revision of logic.

Any conception offering relief from these perplexities must be welcomed and given serious consideration. One such is the 'non-translational' semantics of R. M. Martin's *Truth and Denotation*, whose historical antecedents reach back at least to the theory of semantic inclusion and exclusion outlined by Leibniz in his elegant little *Definitiones Logicae*.[1]

---

[1] Concentrated in three pages as § 13 of G. H. R. Parkinson's *Leibniz: Logical Papers*. The material part of the theory Leibniz partly axiomatizes here may be gleaned from other papers in this collection, especially §4 and §7.

In translational semantics, the subject matter of a given language *L* is accessible also to its metalanguage. This seems inevitable for semantic theories based on concepts of "extent" like satisfaction, denotation, and reference, which correlate terms with the objects falling under them. The plan of a non-translational semantics is to render part or all of the subject matter of *L* inaccessible to its metalanguage, by reverting to concepts of "comparative extent" like semantic inclusion, exclusion and so forth, which correlate terms with one another. Martin bases his non-translational systems on the concept of "comprehension" holding between terms like 'quadruped' and 'fox' whose extensions include one another; it is the extensional counterpart of Leibniz's relation of semantic inclusion.[1] Languages described on the basis of concepts of comparative extent are described more abstractly; the comprehension relation for example does not in general specify the extensions of the predicate constants which it relates. This has the interesting consequence that the semantic theory makes no special commitment to the ontology of the object-language *L*. In a sense it renders the object-language itself free from any specific ontology.

According to Martin, the semantic paradoxes can be circumvented by the employment of non-translational metalanguages. More particularly he sketches a system "SSMc" for which he claims:

> Not only [is it] consistent relative to elementary syntax, as we have seen; it can also be formulated so as to contain its own semantical truth-concept (*op. cit.*, p. 223).

The possibility of the construction of such a system is explained in this way:

> That the truth-concept for SSMc is definable within itself might appear to contradict the result of Tarski mentioned above that, roughly speaking, the semantical truth-concept of a language is not definable within that language, if the language is consistent. But Tarski's result is concerned exclusively with systems of translational semantics (p. 225).

These striking claims embody a novel diagnosis of the major conceptual barriers to the construction of semantically closed languages. They suggest that the expressive capacity of the metalanguage might be materially enhanced through a kind of disciplined renunciation in its lower reaches.

I propose to show to the contrary that no such advantage inheres in the non-translational approach. Beginning with the system SSMc as a case study, it will be shown that it does not after all express its own concept of truth. More generally it will be shown to be incapable in principle of doing so, belonging as it does to a type of language which cannot be semantically closed. The argument proceeds by finding that non-translational systems are indeed subject to semantic paradoxes, which can arise in such a manner as to jeopardize the adequacy of those systems and to bring them under Tarski's limitative results. This means finally that the paradoxes are independent of the issue of the translational or non-translational character of the metalanguage. Put thus positively, the analysis may help to bring into sharper perspective the actual differences among alternative approaches to semantics.

---

[1] To obtain the extensional interpretation of Leibniz's theory, replace 'includes' by its converse throughout the axioms.

The contrast between translational and non-translational semantics is explained in this fashion:

> The various semantical meta-languages developed in the preceding chapters have one important feature in common, namely, that they contain in one way or another the whole of the object-language *L* to which they are applied. . . . We may speak of any such semantical meta-language as a translational meta-language and the semantics involved as a translational semantics (*op. cit.*, p. 179).

On this criterion, any language capable of expressing its own semantic theory seems bound to be at least partly translational, which already suggests the possible emergence of semantic paradoxes somewhere in the upper reaches of the system.

Among non-translational metalanguages therefore it may be crucial to distinguish two grades: the pure type which are *nil*-translational and the mixed type which are *semi*-translational. Both types face a certain problem in connection with Tarski's well-known condition of adequacy:

> One of the important features of non-translational semantics is that the requirement of adequacy for truth-definitions must be abandoned. The reason is that it cannot even be stated for the formalism SMLC, because SMLC contains only structural descriptions, not translations, of the expressions of *L* (p. 184).

The problem is that some or all of Tarski's "T-equivalences" (instances of the schema: *k is true iff P*) may be wholly inexpressible in such systems, and thereby not available as guides to the construction of an adequate truth-predicate. By the same token, they will also be unavailable as premisses in the derivation of semantic antinomies. According to Lesniewski's diagnosis,[1] the Liar and allied antinomies essentially depend on such premisses. Thus it may seem that nil-translational systems at any rate would be safe from this particular source of contradiction.

Concerning his nil-translational system SMLC, Martin claims:

> It is not that adequacy is in any way denied here or its negative assumed. That the truth-concepts defined are clearly in accord with it cannot be explicitly proved, but must rest rather on intuitive grounds concerning the precise meaning of 'Cmprh' (p. 184).

---

[1] See note 7 of Tarski's 'Semantic Conception of Truth' for this attribution to Lesniewski.

Concerning semi-translational systems like SSMc, the situation is fundamentally different. Here at least a partial condition of adequacy can be formulated to control semantic investigation:

> *Condition* T^*: Relative to a chosen translation-scheme, a predicate T of a semantic theory expresses the concept of truth for a language L only if no T-equivalence for L has a translation whose negation holds in that theory.[1]

Condition T^* expresses the comparatively weak requirement that it should be possible to embed the semantic theory in some consistent theory in which T would be an adequate truth-predicate for L in substantially Tarski's original sense. It will be demonstrated that the system SSMc falls short of the possibility of meeting Condition T^*, so that the claim cannot be upheld that 'there is no real conflict here with Tarski's result' (*op. cit.*, p. 226).

Martin's system SSMc is based on a single semantic primitive 'Cmprh' designed to hold between two predicate constants whenever the extension of the first includes that of the second. A predicate constant for Martin is any abstract constructed from a formula having at most one free variable; and the axioms provide for a boolean algebra of predicate constants under the primitive semantic relation. Every expression of the language is provided with a structural description within the language, which also has variables ranging over all of its own expressions. For this system Martin proposes a truth-definition which can be glossed:

> *Definition V* [*TD* 119]: 'Tr *a*' *for* '*a* is a sentence, and the predicate constant formed from *a* by abstraction comprehends all predicate constants'.

Now an argument will be offered to show that 'Tr' so defined does not express the concept of truth for SSMc. This argument will exemplify some more general features of self-applied non-translational languages.

In the first place these systems are liable to emerge as semi-translational in admitting of the definition of at least a partial denotation concept (Theorem 1). This brings them under the condition of adequacy T^* and renders them vulnerable to Grelling's paradox (Theorem 2). The way in which this paradox arises is particularly instructive in that it will not necessarily lead to an actual contradiction within the system. This is due to the liberality of our partial adequacy condition T^*. It may be possible to derive an instance of the schema *k is false iff P* even though the system is too weak to allow a derivation of the correlative *T*-equivalence [*k is true iff P*] which would be needed to obtain an actual contradiction. Indeed I shall not controvert Martin's claim that SSMc is a consistent system (p. 225), but only his claim that it is adequate to the purposes for which it was developed. Having argued that Definition V does not capture the concept of truth for SSMc, I will go on to show that the concept in question is altogether beyond the expressive capacity of that system (Theorem 3), and finally will draw out some consequences of this more general expressive limitation.

---

[1] Quine's noncommittal term 'holds' is used here to accomodate theories of various degrees of formalization (*from a logical point of view*, p. 134). For the present analysis, the relevant interpretation of 'holds in the theory' is 'derivable as a theorem'.

Initially we find that the concept of multiple denotation (pp. 99f.) with its right-field restricted to linguistic expressions, is definable within SSMc:

> *Theorem* 1: *The system SSMc contains the means for expressing its own concept of metalinguistic denotation.*

The simplest definition would identify '*a* denotes *b*' with '*a* comprehends the structural description of *b*'; but for present purposes a more complex definition leads more directly to the end in view:

> *Definition* 1: '*a* denotes *b*' abbreviates 'Tr ϕ' (where *a* is any predicate constant formed from any sentential function *d* of one free variable *c*, and ϕ is that sentence which results when each free occurrence of *c* in the sentential function *d* is replaced by the structural description of *b*).

For this style of definition, compare Martin, *op. cit.*, p. 219. We may suppose an appropriate definition of the notion of a structural description within the elementary syntax of SSMc. Given requisite modifications on the semantic rules to accord structural descriptions the status of complex individual constants, the equivalence of our two definitions can be shown by way of a modified version of Theorem TE6a (P. 222). A typical case of Definition 1 might be glossed: *'Logical Constant' denotes Tilde* abbreviates *'Tilde is a logical constant' is true*, on the temporary hypothesis that 'Tr' expresses the concept of truth.

Now a sequence of abbreviatory definitions will be given for certain structural descriptions of the system: the structural description of a particular sentential function, predicate constant, singular term, and sentence, respectively:

> *Definition* 2: '*m*' abbreviates the structural description of '$\sim$ (*a* Den *a*)'.

Definition 3: 'h' abbreviates the structural description of the predicate constant which results from binding the free variable of m with the abstraction operator.

Definition 4: 'H' abbreviates the structural description of 'h'.

Definition 5: 'G' abbreviates the structural description of 'h Den h'.

The predicate constant h is of particular interest in that it expresses the concept of heterologicality, on the hypothesis that 'Tr' expresses the concept of truth. In Martin's notation, H is 'ay$\cap$invep$\cap m$', and h is '$a\ni \sim (a \text{ Den } a)$'. We now obtain:

Theorem 2: *The predicate 'Tr' constructed according to Martin's Definition V does not express the concept of truth for SSMc, in the sense that it violates Condition* T*.

One theorem from Martin is needed, his TD4, to the effect that any sentence a is false if and only if its negation (tilde$\cap a$) is true (p. 197). From Definitions 1 to 5 we obtain as a theorem of SSMc:

$$\mathrm{Tr}(\text{tilde}\cap \mathrm{G}) \equiv \mathrm{h \ Den \ h}$$

which by Martin's TD4 yields

$$\mathrm{G \ is \ false} \equiv \mathrm{h \ Den \ h}.$$

More generally, the flaw in SSMc which Theorem 2 exposes depends on no peculiar features of Martin's definition of 'Tr' but is endemic to the system, as a simple argument shows; for no predicate constant of SSMc however defined can express the concept of truth for SSMc. This is most readily demonstrated by way of a certain adaptation of Tarski's definability theorem due to R. Smullyan:[1]

Definability Theorem (*Tarski-Smullyan*): *For any "normal" semantic system, the complement of the set of true sentences is not definable within the system.*

Any predicate constant *defines* the set of all those expressions it denotes; as to normality, the *norm* of any expression is that expression followed by its own structural description. For example, the norm of the predicate constant h is the sentence $h\cap H$, which purports to express the proposition that 'heterological' is heterological.

Smullyan calls a semantic system normal if the system has a structure which guarantees that a set *W* is definable in the system only if the set $N(W)$, of all those expressions whose norm is in *W*, is also definable in the system. That SSMc is semantically normal can be shown by way of the schematic definition:

---

[1] R. M. Smullyan, 'Languages in which Self-Reference is Possible', *Journal of Symbolic Logic* 22 (1957).

> *Definition* 6: For each predicate constant *q*, let '*nq*' abbreviate the predicate constant 'has a norm which *q* denotes'

where 'norm' itself is defined within SSMc in an obvious way: ('*a* norm *b*' abbreviates '$(\exists w)(w$ StrDes b & $a = b \cap w)$'). For any predicate constant *q* the above schema provides a definition of another predicate constant *nq* related to the first in the requisite way: if *q* defines the set *W*, then *nq* defines the set $N(W)$. Therefore SSMc is a normal system, and the definability theorem precludes the possibility that the complement of the true sentences of the system can be defined within the system. Furthermore, SSMc is what Smullyan calls a 'complemented' system: it has a structure which guarantees that a set *W* is definable in the system only if the set of all expressions *not* in *W* is also therein definable. That SSMc is complemented is ensured by Martin's Theorem T4A, to the effect that every predicate constant of the system has a "negative" in the system (p. 192). Therefore:

> *Theorem* 3. *The set of true sentences of SSMc is not definable within SSMc.*

Being both normal and complemented, the system is inherently incapable of expressing its own concept of truth, quite independently of the question of its logical consistency. Each predicate constant of SSMc either comprehends the structural descriptions of some non-truths or it fails to comprehend the structural descriptions of some truths.

It is an open question whether the style of comprehensional semantics can be salvaged in a theory capable of expressing its own concept of truth. Minimally, the axioms of comprehension require modification in some fashion so as to render the system either non-normal or non-complemented. Of the two conditions, complementation seems the more readily dispatched, particularly in application to languages admitting of restrictions on types and categories of terms. The development of such a theory along with appropriately strengthened conditions of adequacy may still hold considerable promise in application to the semantics of natural languages, with their intricate category-structure, where closure principles like the "law of semantical negatives" may indeed fail. And it would have intrinsic philosophical interest for many of the reasons outlined in the last chapter of *Truth and Denotation.*

It is intriguing to note in passing that Leibniz's theory does indeed repudiate complementation in favour of a second primitive of comparative extent, the semantic exclusion relation. Leibniz gives a number of axioms to govern semantic exclusion in its interaction with semantic inclusion, the whole enterprise being informed by the desire 'to do without [negative] terms',[1] which is to say, to explain the phenomena of grammatically negative sentences wholly in terms of semantic exclusion.

But these considerations of complementation and the like obtain equally for translational as well as non-translational systems, which leads into the general conclusion to be drawn from the examination of the system SSMc. Indications are that the joint problem of resolving the paradoxes and conceptualizing semantically closed languages[2] is altogether independent of the difference between translational and non-translational metalanguages. Whatever special devices or conceptual reforms are suited to this undertaking are likely to be adaptable to either form of semantic theory and seem indispensable for both.

---

[1] From 'General Inquiries about the Analysis of Concepts and of Truth', §7 of *Leibniz: Logical Papers* (p. 67).

[2] Current developments in this direction are discussed in *The Liar Paradox* (Robert Martin and B. Skyrms, eds.), Yale Press, forthcoming.

University of Toronto

POPPER'S VERISIMILITUDE

By G. S. Robinson

AS long ago as 1960 Popper[1] proposed a definition of 'verisimilitude' to provide a sense for the notion of scientific progress, a notion which is otherwise a vague, impressionistic one, one, it has even been suggested, that reflects no more than the changes in the standards against which we measure the achievements of the past. In complete opposition to this latter point of view, Popper attempted to provide a test or measure that was 'objective', by this meaning a test that was not a function of, and did not reflect, the point of view of any individual or group. It is not clear whether he thought he could eliminate entirely any need for judgment by individuals or groups in the application of his test, but his aim seems to have been to reduce it to a minimum by defining 'verisimilitude' in terms of 'truth', a notion which can be made to look objective and quite independent of the judgment of individuals or groups—unless the group happens to be the whole of a linguistic group, or humanity. Together with the notion of truth, which may or may not be 'objective' in the required sense, Popper's definition purported to involve only the mechanical application of techniques of logic and mathematics, which are for many people the very ideals and models of objectivity.

On Popper's proposed definition, the verisimilitude of a statement or theory consists in the excess of the number of true logical consequences of that statement or theory over the number of false logical consequences. This would allow us, he felt, to make comparisons between theories that are known to be false and to say, for instance, that Newton's dynamics was an advance and represented progress compared to Kepler's or Galileo's theories 'even though we may regard it as refuted'[2]. It had greater verisimilitude in Popper's sense.

Of his conception of verisimilitude Popper seems to believe (1) that it is objective; (2) that it is applicable to scientific theories; and (3) that it is consonant with his general methodological views. On all three points he is wrong, and since this seems not to have been noticed, it is perhaps worth pointing out even though his concept of verisimilitude has not had a large impact on the understanding or practice of science.

Because the formal definition of verisimilitude and the concepts it involves is inapplicable in practice it in fact decreases genuine objectivity by concealing the point at which judgment enters under a cloak of formal procedures, and allows intuitive assessment to masquerade as the outcome of logical and mathematical calculation. This sort of situation is all too familiar in these days when so much work, especially in the social sciences, consists in the mistaken application of mathematical and logical formalisms to inappropriate materials, giving a wholly false sense of accuracy and objectivity. If scientists were to take Popper's conception of verisimilitude and progress seriously it would have the effect of stultifying growth and progress because what he calls 'verisimilitude' and 'progress' could be increased or even maximized by a policy of incurious repetition of safe experiments.

---

[1] K. R. Popper, *Conjectures and Refutations*, Ch. 10, sec. IX–XIII, plus addendum 3.

[2] *Ibid* p. 236.

The key notion of Popper's definition is that of the 'content' of a statement or theory. This consists of 'the class of all the logical consequences'[1] of the statement, whether true or false. One is supposed to divide this, supposedly infinite, set into the true consequences of the statement or theory and its false consequences, discarding anything which is itself derived from one of the false consequences. Popper then adds the disarming sentence: 'If we now work with the (perhaps fictitious) assumption that the content and the truth-content are in principle *measurable*, then we can go beyond this definition and define Vs(a), that is to say, a measure of the *verisimilitude* or *truthlikeness* of *a*.'[2] But here, if not earlier, the whole project has entered the realm of fantasy where it may give an illusion of the possibility of measurement and precision that corresponds to nothing practical whatever. The illusion is helped along by words such as 'measurement', 'logical consequence' and a certain amount of logical formalism, all of which are thought of as having a precise use and meaning. But the notion of logical consequence is *not* here being used precisely. Single statements, certainly the kind that are found in, or that express, theories, do not by themselves have *any* logical consequences that are true or false. If we allowed ourselves to add to them statements of initial conditions we will generally be favoured with as many consequences as initial condition statements, and most of them will probably turn out to be false, taken in any strict logical sense, because the precision allowed by laboratory techniques hardly matches up to what can be achieved by mathematical manipulation. But if we put aside these practical matters of *telling* whether a consequence is true or false or even whether the initial condition statement has been satisfied, and stay firmly in a rather disembodied, theoretical realm where consequences are simply true or false without further ado, we are still not out of difficulty. Abstractly considered, the theoretical statements together with initial condition statements will generally produce an infinite number of consequences, and in this situation there will be only two degrees of verisimilitude:- maximum and nil according to whether the 'truth-content' is infinite and the 'falsity-content' finite or whether they are both infinite. (Though it is of no practical account, I suppose one should add, in order to satisfy those with a taste for theoretical completeness, that the 'falsity-content' might be infinite and the 'truth-content' finite and also, I suppose, that one might be denumerably and the other indenumerably infinite, *etc.*, *etc.*] with increasing degrees of irrelevance.)

---

[1] *Ibid* p. 233.
[2] *Ibid* p. 234.

Finding this abstract and rarified realm uncongenial for Popper's notion of 'content' and 'verisimilitude' we might try to avoid the difficulties by descending into the real world where actual scientists decide and discover things. By limiting the 'truth-content' and the 'falsity-content' of theoretical statements to those consequences that are known or believed to be true and those which are known or believed to be false, we can at least restrict them to a finite number even though we may introduce a regrettable element of judgment that reduces the 'objectivity' of our measure. But if we do bring the concept into some sort of relation with actual scientific practice in this way, we will find that it would have a disastrous effect if scientists were to treat this sort of 'verisimilitude' as something to be aimed at and as constituting scientific progress. If the verisimilitude of any theory or statement were measured simply by the number of known true consequences less the number of known false consequences a scientist could increase the verisimilitude of his theories as much as he pleased, and supposedly advance science, by mindlessly repeating safe, confirming experiments.

Now this sort of result ought not to surprise anyone. If we wrongheadedly seek formal and mechanical definitions of such notions as progress or scientific advance we will inevitably be led into difficulties of just this kind. Human activities of the scope and scale of science are not to be conducted or assessed by means of algorithms or decision procedures. Popper appears to manage because he keeps his test firmly out of contact with practical details and because he is seeking a test that will confirm comfortably our intuitive judgments, *e.g.* that Newton represents an advance over Galileo, (surprise!). If we try to apply his test to cases a little less straightforward than that we will find it useless, or dangerous, or both.

University of Southampton

THE INFALLIBILITY OF OUR KNOWLEDGE OF OUR OWN BELIEFS

By U. T. PLACE

IT would seem to be the case that if I assert a proposition of the form 'I believe that *p*', I cannot be mistaken. The proposition 'I believe that *p*' may be false; but if it is, it must be the case that I am lying. Why should this be?

On what Ryle, in *The Concept of Mind*, called 'the official doctrine', the infallibility of my knowledge of my own beliefs is explained on the assumption that a belief is a private inner state that I and only I can inspect. The objection to this theory is that if it were true, our knowledge of our own beliefs would be incorrigible in the sense that no one else is able to inspect my beliefs and thus correct any error I may have made in reporting them; but it would not be infallible, because there is no reason in principle why I should not make a mistake through careless introspection even though I am the only person who is in a position to correct it.

A more promising answer to the problem is suggested by Ryle's dispositional theory of mental concepts. It might be argued on this theory that to believe that *p* is to be disposed to assert *p* on occasions when the truth of *p* is a relevant consideration. Now since I cannot assert that I believe *p* without *ipso facto* asserting *p*, it would follow that in asserting that I believe *p*, I have *ipso facto* exercised and thus displayed my disposition to assert *p*. On this view, I cannot be mistaken in asserting that I believe *p*, because the statement 'I believe *p*' is a self-verifying statement. It cannot be asserted without *ipso facto* demonstrating its own truth.

There is, however, a formidable objection to this explanation of the infallibility of our knowledge of our own beliefs in that it does not allow any room for the case where my assertion that I believe *p* is a lie. For if to believe that *p* is merely to be disposed to assert *p*, I cannot assert that I believe *p* without demonstrating that I believe *p*. The statement can never be false. Consequently, in order to allow for the case of lying, we have to amend the dispositional theory of belief by adding to the disposition to assert *p* the disposition to act on *p* in circumstances where *p* is or would be a relevant consideration. It may be objected that an account of belief in terms of the disposition to act on a proposition one is disposed to assert, runs into difficulties in the case of beliefs that are related to matters far removed from the believer's immediate life situation, such as the non-historian's beliefs about the remote past or the non-astronomer's beliefs about the heavenly bodies.[1] However, since no one could conceivably have a motive for lying in claiming to believe something about a matter of no immediate concern to himself, the problem of distinguishing between lying about one's beliefs and sincerely asserting them does not arise with beliefs of this type. Hence there is no objection to supposing that where the belief concerns matters outside the believer's immediate life situation, his belief consists solely in a disposition to assert the proposition he believes and that it is only in the case of beliefs that relate to the individual's immediate life situation that believing *p* also involves a disposition to act on *p*.

On this amended form of the dispositional theory, the case where I sincerely assert that I believe *p* is a case where I am disposed both to assert and to act on *p*, whereas in the case where I am lying in asserting that I believe *p*, I am disposed to assert *p*, but am not disposed to act on *p*. On this amended view, the statement 'I believe that *p*' is still partially self-verifying in that it demonstrates the existence of the disposition to assert *p* which is part of what is meant by saying that I believe *p*. On the other hand it is not self-verifying with respect to the disposition to act on *p* which, as we now construe it, is also implied by the statement that someone believes that *p*. But if I do not show that I am disposed to act on *p*, when I claim to believe *p* in the way that I show that I am disposed to assert *p*, how does it come about that I can lie, but cannot be mistaken, when I falsely assert that I believe that *p*? Why should I not sincerely assert that I believe *p*, and yet be honestly mistaken in supposing that I am disposed to act on *p*?

It is no use saying here that if I assert that I believe *p* in a case where I am not disposed to act on *p*, I must necessarily be telling a lie. This is merely to sweep the problem under the carpet. For to lie is to assert a proposition that one knows to be false. Hence a situation in which any false proposition must necessarily be a lie, is merely a situation in which the speaker cannot avoid knowing that the proposition is false, if it is false. If, therefore, I am necessarily lying, in asserting that I believe p, when I am not in fact disposed to act on *p*, it follows that in asserting that I believe *p*, I cannot avoid knowing that I am not disposed to act on *p*, if I am not so disposed. But why should I necessarily know that I am not disposed to act on *p*, when I assert that I believe *p*, but am not in fact disposed to act on *p*? What is the force of 'necessarily' here? Is this a matter of logical necessity? Or is it merely a contingent psychological necessity?

I suggest that in order to answer this question we need to consider what is involved in asserting a proposition. Clearly, if I assert a proposition *p*, I am performing a social act, an act that presupposes not only a speaker, but also an audience. Moreover, my intention in asserting *p* must necessarily be to influence my audience in the direction of believing *p*. If I do not have this intention, I cannot properly be said to have asserted *p*. But if the account we have given of what is involved in believing *p* is correct, it follows that in trying to influence my audience to believe p, I am trying to influence them in the direction not only of asserting *p* themselves, but also in the direction of acting on *p*. Thus, in asserting *p*, I am necessarily recommending *p* to my audience as a reliable basis for action.

---

[1] I am indebted to Professor P. T. Geach and Mr. W. J. Rees for drawing my attention to this objection.

However, from the fact that in asserting *p* I necessarily intend to influence my audience to adopt *p* as a basis for action, it does not follow that my audience cannot avoid being influenced in this way. Nevertheless, there is a substantial body of empirical evidence relating to the phenomena of suggestion and hypnosis which supports the view that the initial reaction of every human being who hears and understands a proposition asserted by another person is to accept that proposition at its face value as a reliable basis for action, and that the ability to question the propositions that are put to him by another person, rejecting some, while accepting others, is a skill that is superimposed on this basic tendency to accept what one is told without question. Thus a study by Messerschmidt quoted by Hull[1] shows that the responsiveness of children to verbal suggestions of postural movement increases rapidly as the child's understanding of language increases up to an average age of 8 years and then slowly declines. Hull comments as follows:

> As a plausible hypothesis to account for this reversal, it may be supposed that suggestion is based in a primitive habit tendency (of responding directly to verbal stimulation) which is useful in most situations but maladaptive in the special type of situation represented by this suggestion test. Presumably the maladaptivity is related largely to the fact that if a person responds positively and indiscriminately to all suggestions made by others, he is likely to be taken advantage of by his associates in that the energies needed for his own welfare will be diverted to that of those giving the suggestions. The rise of the curve accordingly represents the acquisition of a working knowledge of the language, which obviously must proceed a certain distance before its maladaptive possibilities may be encountered; and the gradual fall observed from about eight years on may be regarded as an indication of the progress in 'unlearning' those particular reactions to verbal stimuli which, having been established have proved maladaptive (p. 85).

In the light of empirical evidence such as this, as well as in the light of a theoretical consideration of what is required for a child's acquisition of the ability to understand what is said to him, it seems not unreasonable to suppose that the child must necessarily begin by learning to accept a statement made by another person as equivalent in all respects to the actual existence of the environmental situation which the statement describes, and that unless he begins in this way he cannot learn to understand the meaning of what is said to him. In other words, unless a child begins by accepting whatever he is told and believes it implicity, he cannot learn to understand what is said to him. If this is correct, it follows that anyone who listens to and understands an assertion made by someone else will necessarily believe that assertion and thus be disposed to act upon it, unless he has acquired an overriding disposition to reject assertions of that kind.

---

[1] C. L. Hull, *Hypnosis and Suggestibility*, New York, Appleton-Century, 1933.

Now, in learning to reject certain assertions while admitting others, the individual must have available to him certain cues which enable him to differentiate between acceptable and unacceptable assertions made by others. It seems, moreover, that the cues to which he learns to respond in this way are of three kinds. In the first place he may learn to reject out of hand, or not to admit without serious question, propositions which are asserted by certain persons whom he has learned to identify as persistently unreliable informants or inveterate liars. Secondly, he may learn to reject or seriously question propositions asserted by persons whose propositions he would otherwise accept at their face value in a situation where the person concerned is judged to have a strong motive for lying or otherwise misrepresenting the true state of affairs. Finally, he will also learn to question and ultimately reject certain propositions purely on their merits, regardless of the source from which they come. When he does this, he will do so on the basis of some contradiction that he notices or suspects between the proposition that is being asserted and certain other propositions which he already believes and has successfully acted upon in the past. It would seem to be the case, moreover, that if he has no reason to question the veracity of his source and is not aware of and does not suspect any conflict between his existing beliefs and the proposition being urged upon him, there is nothing to prevent him from giving way to this underlying tendency to accept what he is told as a basis for subsequent action.

Now if, as I have argued, a human being has an in-built tendency to accept as a basis for action any proposition that is asserted by another person which he hears, or reads, and understands, unless there is something that restrains him from doing so, it seems not unreasonable to suppose that he will likewise have a built-in tendency to accept as a basis for action any proposition which he himself asserts. For no one can effectively assert a proposition without hearing and understanding what he is saying or without reading and understanding what he is writing. Hence on the principle whereby the inveterate liar eventually comes to believe his own propaganda,[1] an individual who asserts a proposition will not be able to resist the temptatio nto take the proposition he asserts as a reliable basis for action, unless there is something to prevent him from doing so. Now we have seen that the only conditions under which an individual can be restrained from accepting a proposition that he hears or reads and understands, are (a) in a case where he has learned that the speaker is, or is likely to be, a systematically unreliable informant, (b) in a case where he has learned that the speaker is likely to have motives for mendacity or misrepresentation and (c) in a case where he is aware of or suspects a conflict between the assertion and his existing beliefs. Applying this to the case where the individual understands the proposition he himself asserts, as he must do whenever he asserts it, it is evident that no rational being could learn to treat himself as a systematically unreliable informant without systematically undermining all rational connection between what he says, whether to himself or to others, and what he does. That way lies madness.

---

[1] I am indebted to Professor P. T. Geach for drawing my attention to this principle and for his criticisms of an earlier draft of this section of my paper.

On the other hand, he can very profitably learn to question, though not systematically to reject, those propositions that he asserts or is inclined to assert where he himself has strong motives for mendacity or misrepresentation. This is a highly desirable habit of self-criticism which few people manage to acquire in the face of the strong impulse that we all have to believe, whenever possible, what it suits us to believe, rather than what the evidence demands. So difficult is it to learn this discrimination, so directly does it conflict with the individual's natural inclinations as far as the beliefs he accepts and rejects are concerned, that it is inconceivable that a man who for this reason does not accept as a basis for action a proposition he has just asserted should be unaware that he does not in fact believe the proposition in question.

The only other case in which he could reject as a basis for action a proposition which he himself has just asserted is a case in which he is aware of or suspects a conflict between the proposition he has just asserted and other beliefs of his that he accepts as a basis for action. In this case, if, as I have argued, he is aware of, or suspects, a conflict between the proposition he has asserted and his other beliefs, it is again inconceivable that he should fail to be aware that he has been led by this apparent conflict to reject the proposition in question as a basis for action. But if in the only two cases where a rational man can reject as a basis for action, a proposition he has just asserted, it is inconceivable that he should be unaware that he has so rejected it, it follows that it is inconceivable that he should not know that he does not believe the proposition he has asserted, if he does not. And if he cannot avoid knowing that he does not believe a proposition he has asserted, if he does not, it follows that in such a case he must necessarily be lying if he claims to believe the proposition he asserts.

It will be noted that on this account not only is it inconceivable that an individual should not know that he does not believe a proposition he asserts, if he does not believe it, it is equally inconceivable that he should not know that he believes a proposition which he has considered, has accepted as a basis for action, but denies that he believes. For, as we have seen, someone who hears or reads and thereby understands a proposition, as he must have done if he has considered it, cannot fail to be influenced in the direction of accepting it as a basis for action, unless he consciously rejects it either on the grounds of the probable mendacity of its source, or on the ground of a conflict with his existing beliefs. But since he cannot fail to know that he has consciously rejected it, if he has, it follows that he cannot fail to know that he believes it, if he has considered it and has not consciously rejected it. Hence in denying that he believes a proposition that he has not consciously rejected, he necessarily knows that he does in fact believe the proposition in question in which case he is lying in denying that he believes it.

On the other hand, the account I have given does appear at first sight to allow that someone might either believe or not believe a proposition and yet not know that he believes it or does not believe it, provided that the proposition in question is one that he has not yet asserted or considered; since if he has not either asserted or considered it, he would not have any grounds for knowing whether or not be believed it. His situation would be like that of the woman who said she would not know what she thought about the question at issue until she heard what she had to say about it.

However, the implication in the case of the woman who did not know what she believed until she heard herself speak is that she has not yet made up her mind on the issue and only does so as she speaks. In other words, she does not yet either believe or not believe the propositions she will come to believe when she asserts them. But the case we are considering is one in which someone already believes something, but does not know that he believes it because he has not yet asserted or considered it. The existence of such a case is in fact ruled out on the account of belief that I have suggested, since the only way in which an individual can acquire a disposition either to act or not to act on a proposition is as a result of hearing or reading and thereby understanding the proposition in question. If he has never asserted or considered it, he cannot have heard, read or understood it, and cannot, therefore, have a disposition to act on it; and if he does not have a disposition to act on it he cannot be said to believe it.

These considerations do not rule out the possibility that someone who has never in fact considered a given proposition or, if he has, has rejected it, might nevertheless be disposed to act *as if* he believed the proposition in question. For example, someone who suffers from an irrational phobia for cats is disposed to act as if he believes the proposition 'Cats are dangerous', although he would quite properly deny believing this proposition. In such a case, his disposition to behave in this way does not proceed from his acceptance of the proposition he denies believing. It is an irrational impulse that exists independently of and despite his beliefs. Nevertheless, although the cat phobic can quite properly claim not to believe the proposition 'Cats are dangerous' since he is not in fact disposed to assert this proposition, he cannot claim that he believes the contrary proposition 'Cats are not dangerous'. For on the view I am defending, to believe that cats are not dangerous entails being disposed not only to assert that they are not, as the cat phobic usually is, but also being disposed to act on this proposition which the cat phobic is manifestly not disposed to do.

Thus we cannot say of the cat phobic either that he believes that cats are dangerous, or that he believes that they are not. The trouble is that he does not behave like the man who cannot make up his mind whether to believe *p* or not-*p* and hence does not believe either. The phobic behaves like a man who *has* made up his mind and has made up his mind that cats are dangerous. Consequently, both the phobic himself and those who observe his behaviour find themselves in some conceptual perplexity when they attempt to characterise his state of mind. Going on his behaviour, we are inclined to say in spite of his protestations to the contrary that he shows quite unambiguously that he believes cats to be dangerous. But if he really believes that cats are dangerous, then his denial that he believes they are dangerous must be false and he must either be lying or be mistaken. Yet in such cases it is usually quite evident not only that the phobic is perfectly sincere in denying that he believes cats are dangerous, but also that he is in no way deceived as to the character of his own mental and behavioural dispositions.

The truth is that in denying that he believes cats to be dangerous, he is not denying for one moment that he has an overpowering disposition to act as if they were. All he is denying is that this disposition proceeds from an acceptance of the proposition 'Cats are dangerous'. It is true that his acceptance of this proposition at some time in the past may have contributed to his present disposition to avoid cats. But if so, it is equally true that it is no longer sustained by any such acceptance. This is shown by the fact that however often he may rehearse to himself the contrary proposition 'Cats are not dangerous' together with all the arguments and evidence that can be adduced for this proposition and against the proposition 'Cats are dangerous', he cannot induce in himself a disposition to act on the proposition 'Cats are not dangerous' in the face of the overwhelming irrational impulse to act as if they were.

Thus the case of the irrational phobia is not an example of someone who is mistaken in his assessment of his own beliefs; it is an example of someone in whom there is a disturbance of the normal and 'rational' relationship between what a man asserts and the way he behaves, which is implied by the concept of belief.

If this account, or something like it, is correct, it is apparent that no definitive answer can be given in the present state of knowledge to the question 'Is the necessity that makes it impossible to avoid knowing that we believe a proposition, if we do, or that we do not believe it, if we do not, a logical necessity or a contingent psychological necessity?' If it turns out, as it may, that it is logically impossible to conceive of a language learning process which does not involve the initial acquisition of the tendency to accept at its face value as a basis for action any proposition that is understood, we should then, I suggest, be perfectly justified in regarding the necessity as logical. If, on the other hand, this turns out to be a purely accidental feature of the way human beings are constructed, we might accept it as a contingent psychological necessity. But in either case, the story that has to be told is too complicated for any short answer to have much value.

University of Leeds

ON NOT BEING SAID TO DO TWO THINGS

By GARETH B. MATTHEWS

> To say that a person has been enjoying digging is not to say that he has been both digging and doing or experiencing something else as a concomitant or effect of the digging; it is to say that he dug with his whole heart in his task, *i.e.* that he dug, wanting to dig and not wanting to do anything else (or nothing) instead. His digging was a propensity-fulfilment. His digging was his pleasure, and not a vehicle of his pleasure.[1]

THE idea expressed in these words of Gilbert Ryle is both traditional and characteristically Rylean. It is traditional insofar as it echoes ideas about pleasure and enjoyment to be found in Aristotle. It is characteristically Rylean in that it exhibits the tactic, familiar to all readers of *The Concept of Mind*, of arguing that where the dualist supposes there are two acts or processes or events—one physical and the other mental—there really is only one. What the dualist takes to be a shadow or mental act, process or event alongside the physical one is really, according to Ryle, something else—a disposition, perhaps, or, in the case of enjoyment, the fulfilment of a disposition.

---

[1] *The Concept of Mind* (London, 1949), p. 108.

Ryle expects to establish his point by appeal to the meaning of what we say when we say that someone is enjoying doing something. His claim is that *to say* that a person has been enjoying digging is *not to say* that a person has been doing two things, or even doing one thing and experiencing another; it is *to say* that he is doing something in a certain way ('with his whole heart in his task').

Can it be that Ryle is here making the jejune point that

(1) Sam enjoyed digging

is not identical with

(2) Sam dug and enjoyed it?

Surely not. Ryle means to be telling us something about what it is to say that someone has been enjoying doing something. If his comments were rendered irrelevant by the innocent switch from (1) to (2), they could be of no philosophical interest. Clearly Ryle's analysis must apply equally to (2). That is, he must want to maintain that to say one has been digging *and* enjoying it is not to say that one has been 'both digging and doing or experiencing something else as a concomitant or effect of the digging'.

So far the point Ryle wants to make parallels his claim that 'when a person is described as having fought and won, or as having journeyed and arrived, he is not being said to have done two things . . . .' (*op. cit.*, p. 150). The conjunction in 'He fought hard and won' and the one in 'She fell down and hurt herself' might each be called a 'conjunction of consequence'. Thus it is not enough to make 'Alice fell down and hurt herself' true, that Alice have fallen down (on some occasion or other) and that Alice also have hurt herself (on some occasion or other). Alice's hurting herself must be the result or consequence of her falling down: She fell down and, in consequence, or as a result, hurt herself.

The sentence,

(3) Sam dug all morning and wore himself out

contains a conjunction of consequence. The idea expressed is that Sam wore himself out as a result of digging. But the 'and' in

(2) Sam dug and enjoyed it

is not a conjunction of consequence. (2) doesn't tell us that Sam enjoyed digging as a result of digging. So nothing parallel to the reasoning Ryle has up his sleeve on the fought-and-won case (to say someone fought and won is not to say he did two things, since winning is merely the upshot of fighting) will apply here.

What Ryle suggests is that (2) means

(4) Sam dug with his whole heart in his task.

Then the reason for saying that (2) does not report Sam as doing two things (or even as doing one thing and undergoing another) is that it reports him as doing one thing in a certain way.

II

Does 'Sam dug with his whole heart in his task' really mean the same thing as 'Sam enjoyed digging'? I think not. But suppose it did. Why should we think this shows, or even tends to show, that to say Sam dug and enjoyed it is not to say Sam did two things?

The idea must be that, if we show a claim that someone x-ed and y-ed is roughly equivalent to a claim that he x-ed in such-and-such a fashion ('dug with his whole heart in his task'), we have shown that the original claim asserts only that someone did one thing. Unfortunately this reasoning is faulty. Consider an example.

Presumably

(5) Hubert rode along the parade route and smiled as he went

asserts that Hubert did two things—rode and smiled. Yet

(6) Hubert rode smilingly along the parade route

is roughly equivalent to (5). The equivalence (or near equivalence) of (5) and (6) hardly shows that to say someone rode and smiled as he went is only to say he did one thing. One could as well argue that it shows that to say someone rode smilingly is to say he did two things—rode and smiled as he went. Similarly, one could as well argue that the (assumed) equivalence of (2) and (4) shows that to say Sam dug with his whole heart in his task is to say Sam did two things.

What would make Ryle's conclusion more appealing is some reason to think that 'Sam dug and enjoyed it' is a non-standard way of putting what could be put standardly as 'Sam dug with his whole heart in his task'.

III

I can think of two ways to argue that 'Sam dug and enjoyed it' is non-standard. One is to argue that it is only by a figure of speech that Sam could be said to *both* dig *and* enjoy digging. The other is to argue that the 'and' of 'Sam dug and enjoyed it' is what we might call an "ampliative conjunction". I shall discuss the second suggestion first

'I do not like thee, Doctor Fell,' goes the nursery rhyme,

> The reason why I cannot tell;
> But this I know, and know full well,
> I do not like thee, Doctor Fell.

The 'and' of 'know, and know full well' is ampliative. Other examples of ampliative conjunction are these:

> Though a month ago he was doing no campaigning, he is now running, and running hard.
>
> You must learn to think, and think clearly.

Presumably a man who is said to run, and run hard, is thereby said to do one thing (in a certain way). Likewise, the reasoning would go, a man who is said to dig and enjoy it (*i.e.*, to dig, and dig with his whole heart in his task) is thereby said to do one thing (in a certain way).

But now what about 'He digs, and digs with his tongue hanging out'? The conjunction is ampliative all right. Yet one could plausibly say that a man who is said to dig, and dig with his tongue hanging out, is thereby said to do two things—*viz.*, to dig and to let his tongue hang out.

It seems that the mere fact that a conjunction is ampliative gives us no good reason to say that someone is only being said to do one thing. So even if the 'and' of 'Sam dug and enjoyed it' is ampliative, that fact does not show that Sam in (2) is only being said to do one thing.

IV

There remains to consider the possibility that it is only by a figure of speech that Sam could be said to dig *and* enjoy it.

If 'Sam dug and enjoyed it' is a figure of speech in the way suggested, it must be a case of hendiadys. Hendiadys is an ancient trope recognized to be common in Biblical Hebrew and in classical Greek and Latin but thought to be rare in English. The dictionary example 'We drink from cups and gold' (meaning, supposedly, 'We drink from golden cups'), is unhelpful as a way of coming to understand what hendiadys is, since it is not really English at all but a word-for-word translation from Vergil. But 'The farmer waits until his pigs are nice and fat' (='nicely, or agreeably, fat') is genuine English and a genuine hendiadys, as is 'He likes the rough and tumble of politics' (= 'the rough tumble of politics').

Perhaps the most common English examples of hendiadys on verb pairs (what 'dug and enjoyed it' is, if it is hendiadys) are those formed with 'try and . . .', *e.g.*, 'Try and hit it' (='Try to hit it'). (I shall ignore the fact that some grammarians consider this hendiadys ungrammatical.) Certainly a man who is said to have to try and hit the bullseye is not thereby said to have to do two things (*viz.*, to try and also to hit the bullseye) but to do one thing (*viz.*, to try to hit the bullseye).

The Greek roots of 'hendiadys'—which mean 'one through, or by means of, two'—suggest that we might test whether a putative case of hendiadys is genuine by asking whether the words or phrases it joins by 'and' name, denote, refer to, or express only one thing, or two different things. But a moment's reflection will reveal that such a test would accredit all pleonastic conjunctions as instances of hendiadys, which is absurd. Thus 'lord and master', 'to have and to hold', 'to let and demise', *etc.* would all, absurdly, count as cases of hendiadys.

What grammarians and lexicographers usually fix on in their discussion of hendiadys is the fact that the notion expressed by one conjunct is meant to qualify the idea expressed by the other. But the criterion suggested by this sort of observation would also be much too hospitable; conjunctions of consequence would yield cases of hendiadys. Thus 'He ran the mile in 4:05 and won the race' would count as hendiadys on the grounds that the idea expressed by 'won the race' really qualifies 'ran the mile in 4:05'—'He ran the mile in 4:05, winning the race'.

An adequate test for hendiadys must depend, I think, upon the fact that simplification fails for one conjunct in the case of a true hendiadys. Thus from 'He ran the mile in 4:05 and won the race' it follows both that he ran the mile in 4:05 and that he won the race. And from 'She is yours to have and to hold' these both follow: 'She is yours to have' and 'She is yours to hold'. But from 'The farmer waits until his pigs are nice and fat' it does not follow that the farmer waits until his pigs are nice (though it does follow that he waits until his pigs are fat). And from 'Ben has to try and hit the bullseye' it doesn't follow that Ben has to hit the bullseye (though it follows that he has to try).

If I am right about this, then 'Sam dug and enjoyed it' is not a case of hendiadys; for from it we may get both 'Sam dug' and 'Sam enjoyed it'. And so it is not just by a figure of speech that Sam may be said to dig *and* enjoy it.

There seems to be no good reason for insisting that a man said to enjoy what he is doing is said to do one thing rather than two. To say Sam has been enjoying digging is, if you like, to say that Sam has been both digging and enjoying what he was doing.

Nothing very exciting follows from this concession—certainly nothing about enjoyment being an effect of digging, or about digging being sometimes a "vehicle of pleasure". Those ideas would need support from elsewhere. But they aren't blocked from the start.

University of Massachusetts